

# Ergodizitätseigenschaften Iterierter Funktionensysteme unter Kontraktions- und Driftbedingungen nur am Rande

Von der Fakultät für Mathematik der Universität Duisburg-Essen

zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

**Joachim Kowol**

aus Beuthen

Referent: Prof. Dr. Ulrich Herkenrath

Korreferent: Prof. Dr. Denis Belomestny

Korreferent: Prof. Dr. Marius Iosifescu

Datum der mündlichen Prüfung : 29. April 2014

## Danksagung

Zum Ende meines Studiums, die Diplomarbeit war bereits abgegeben, drehte sich alles um den bevorstehenden Berufseinstieg und die Richtung die ich einschlagen würde. Ernsthaft über eine Promotion nachzudenken, wäre mir in dieser Situation jedoch nie in den Sinn gekommen.

So gilt mein Dank in erster Linie natürlich Herrn Prof. Dr. U. Herkenrath, für die großartige Möglichkeit und die fortwährende Zuversicht während der Erstellung der vorliegenden Arbeit.

Außerdem gilt mein Dank Herrn Prof. Dr. D. Belomestny und Herrn PD Dr. V. Krätschmer, für viele richtungsweisende Vorschläge, die die Entstehung der Arbeit oftmals entscheidend vorangetrieben haben.

Abschließend will ich noch meiner Familie danken, die zwar nicht nachzuvollziehen vermochte warum ich mich auch am Wochenende mit Mathematik beschäftigen musste, aber zumindest unerschütterlich an den Erfolg der Unternehmung zu glauben schien.

Duisburg, Dezember 2013

# Inhaltsverzeichnis

<b>Summary</b>	<b>4</b>
<b>Zusammenfassung</b>	<b>6</b>
<b>1 Einführung</b>	<b>8</b>
1.1 Grundlagen . . . . .	8
1.2 GARCH - Beispiel . . . . .	11
1.3 Basrak - Beispiel . . . . .	13
1.4 SLARM - Beispiel . . . . .	15
1.5 Iosifescu - Beispiel . . . . .	17
<b>2 AR(1)-Modell mit ARCH(1)-Innovationen</b>	<b>19</b>
2.1 Smoothing bewahrt die Kontraktionsbedingung . . . . .	23
2.1.1 Glättung mittels Geradenstücken . . . . .	26
2.2 Driftbedingung - geometrische Ergodizität . . . . .	30
2.3 Was Glätten tatsächlich bringt . . . . .	40
2.3.1 Parameterbereicherweiterung dank Glättung mit Geradenstücken . . . . .	40
2.3.2 Polynomiale Glättung . . . . .	45
2.4 Existenz von Momenten der stationären Verteilung . . . . .	52
<b>3 Empirische Studie</b>	<b>55</b>
3.1 Simulation AR(1)-ARCH(1) . . . . .	56
3.2 Simulation Smooth-AR(1)-ARCH(1) . . . . .	58
3.3 Simulation Poly-Smooth-AR(1)-ARCH(1) . . . . .	60
3.4 Parameterschätzung . . . . .	63
3.4.1 Güte der Parameterschätzung . . . . .	73
3.4.2 Predict . . . . .	86
<b>4 Anhang</b>	<b>92</b>
4.1 Kontraktionsbedingungen in 1 und 2 Schritt Übergangsfunktionen . . . . .	92
4.1.1 IFS . . . . .	92
4.1.2 SRE . . . . .	94
4.2 Driftbedingung für den geglätteten Prozess . . . . .	97
4.3 Zusammenhang von $\beta$ und $\alpha_1$ . . . . .	103
4.4 Glättungsfunktion am Rand . . . . .	106
4.5 Prozess-Bausteine . . . . .	109
4.6 garch11Fit . . . . .	114
4.7 Erläuterung ar1arch1Fit . . . . .	116
4.8 Residuenanalyse Smooth-AR(1)-ARCH(1) . . . . .	118
4.9 Predict-Poly-Smooth . . . . .	118
4.10 Normalverteilung und Errorfunktion - Fehlerfunktion . . . . .	120
4.11 Streng Stationärer Innovationsprozess . . . . .	122
Ausblick . . . . .	127

# Summary

Autoregressive time series models have been applied extensively in statistics on financial markets during the last decades (see e.g. McNeil and Frey [8], Basrak [2], Tsay [31]). In particular, the so-called AR(1)-ARCH(1)-model (autoregressive of order 1 with ARCH(1)-noise-term) has attracted much attention in the literature (see e.g. Borkovec and Klüppelberg [19], Cline [5], Würtz and Chalabi and Luksan [22]). In order to ensure (asymptotic) stability respectively stationarity of a corresponding stochastic process sufficient conditions on the parameters of the transition function of the time series-model are searched for.

The time series models mentioned above may be conceived as special “iterated function systems” (abbreviated to IFS). Therefore the already existing theory on those IFS may be applied to the time series models under consideration.

Desirable ergodicity properties of IFS with independent, identically distributed innovations are derived under certain contraction conditions on the corresponding transition function. Some general contraction conditions on IFS are discussed here. Depending on the kind of the condition imposed, different convergence properties of the associated Markov chain of an IFS are proved, in particular the existence of a unique stationary distribution. Such contraction conditions sometimes are too restrictive with respect to a desirable transition function. An example is given by AR(1)-time series with ARCH(1)-innovations (see. e.g. BuK [19]).

In the past the transition functions of IFS were defined by one term over the whole state space of the system. Since this approach may be too restrictive, there are attempts for modifications.

In this context Herkenrath [13] proposes a modification of the AR(1)-ARCH(1)-process, which should avoid “unrealistically high” fluctuations and thereby “extreme” values. The modification is performed by “smoothing” the transition function out of a “normal” compact range. Due to a theorem of Feigin and Tweedie [32] desirable properties of the process are assured by suitable contraction- and drift conditions only outside of the normal range, whereas within that range only continuity assumptions are imposed. Those contraction- and drift conditions in turn are achieved by a composition of appropriate smoothing functions on the transition function outside the normal range.

This approach will be extended in my thesis by studying different types of smoothing functions under theoretical and empirical aspects.

The thesis is organized as follows:



- In Chapter 1 the necessary fundamentals are presented, in order to study properties and consequences of smoothing functions. In Sections 1.2 - 1.5 furthermore different variants of the central concept of “contraction condition” are discussed.
- On the basis of those general results in Chapter 2 various smoothing functions are studied and (minimal) sufficient conditions for the desired ergodicity properties are developed (see 2.1.1). The relations of the smoothing functions to contraction conditions and moreover to drift conditions are examined therefore ensuring geometric ergodicity. This context offers the facility to compare the different variants of smoothing functions. Thus in Section 2.3 positive effects of smoothing are presented in the sense of an extension of the range of parameters ensuring geometric ergodicity. Finally implications for other fields of applications are explained.
- In Chapter 3 the empirical basis for the application of smoothing functions is given. First the algorithms for simulation of the different smoothing variants are discussed, and based on that, the estimation of the corresponding parameters is performed. As conclusion of this empirical, computer-supported study an evaluation of the parameter-estimations is presented.
- Although at first glance Chapter 4 may look as a collection of marginal results, it completes the thesis in important subjects. The detailed explanations (e.g. 4.3) or basic respectively additional examples (e.g. 4.5) were only transferred from the central chapters, in order not to overburden them and not to cover the central ideas there. Reviewing those central chapters and for their deeper understanding they are important, which may be obvious by the multitude of references to the appendix there. An impression and simultaneously an outlook on the modified IFS with a strictly stationary innovation process concludes the thesis.

# Zusammenfassung

Zeitreihenmodelle sind in den letzten Jahrzehnten vielfach zur Modellierung von Finanzmarkt-Zeitreihen, z.B. Aktienkursverläufen, eingesetzt worden (siehe z.B. McNeil und Frey [8], Basrak [2], Tsay [31]). Insbesondere erfreut sich das sogenannte AR(1)-ARCH(1)-Modell (autoregressiv von 1. Ordnung mit ARCH(1)-Störterm) einer intensiven Behandlung in der Literatur (siehe z.B. Borkovec und Klüppelberg [19], Cline [5], Würtz und Chalabi und Luksan [22]). Zur Sicherung der (asymptotischen) Stabilität bzw. Stationarität eines so erzeugten Prozesses suchen die Theoretiker nach dafür hinreichenden Parameterkonstellationen an die Übergangsfunktion.

Die oben genannten Zeitreihenmodelle lassen sich auffassen als spezielle Iterierte Funktionensysteme (abgekürzt IFS). Daher ist die für solche Systeme bereits entwickelte Theorie auf die interessierenden Zeitreihenmodelle anwendbar.

Wünschenswerte Ergodizitätseigenschaften von Iterierten Funktionensystemen mit unabhängigen, identisch verteilten Innovationen lassen sich, ganz generell gesagt, unter gewissen Kontraktionsbedingungen an die Übergangsfunktion gewinnen. Einige generelle Kontraktionsbedingungen, die im Laufe der Zeit für IFS entwickelt wurden, werden vorgestellt und zueinander in Beziehung gesetzt. Je nach Art der Kontraktionsbedingung, z.B. verwendeter Norm oder Metrik, lassen sich bestimmte Konvergenzeigenschaften der bei IFS interessierenden Markov-Kette nachweisen, damit verknüpft die Existenz einer eindeutigen stationären Verteilung.

Solche Kontraktionsbedingungen stellen natürlich eine manchmal unwillkommene Einschränkung an den Entwurf einer Übergangsfunktion eines IFS dar. Ein Beispiel dafür bieten etwa AR(1)-Zeitreihen mit ARCH(1)-Innovationen (vgl. z.B. BuK [19]).

Traditionellerweise werden derartige Modelle bzw. Übergangsfunktionen „im Ganzen“ definiert, d.h. über dem ganzen Zustandsraum des Systems. Da die Kontraktionsbedingungen aber in diesem Fall als einschränkend angesehen werden, gibt es Ansätze, diese mit gehörigem Aufwand abzuschwächen.

In diesem Zusammenhang schlägt Herkenrath [13] eine Modifizierung des AR(1)-ARCH(1)-Prozesses vor, die „unrealistisch hohe“ Schwankungen ausschließen und damit extrem hohe und extrem niedrige Werte vermeiden soll. Die Modifikation geschieht durch eine „Glättung“ der Übergangsfunktion außerhalb des „normalen“ Wertebereichs, einem Kompaktum. Wünschenswerte Eigenschaften des Prozesses, eben (asymptotische) Stabilität bzw. Stationarität, können dank eines Satzes von Feigin und Tweedie [32] durch geeignete Kontraktions- und Driftbedingungen nur außerhalb des normalen Wertebereichs gesichert werden, während innerhalb dieses Bereichs nur Stetigkeitsannahmen nötig sind. Diese Kontraktions- bzw. Driftbedingungen wiederum werden durch die Aufeinanderschaltung von geeigneten Glättungsfunktionen auf die Übergangsfunktion außerhalb des normalen Wertebereichs erzielt.

Diesen Ansatz werden wir in der vorliegenden Arbeit ausweiten und verschiedene Ausformungen von Glättungsfunktionen unter theoretischen wie auch empirischen Aspekten untersuchen.

Die Arbeit ist wie folgt gegliedert:

- In Kapitel 1 werden zunächst die notwendigen Grundlagen geschaffen und Rahmenbedingungen umrissen, um Eigenschaften und Effekte von Glättungsfunktionen zu untersuchen. Anschließend werden in den Abschnitten 1.2 - 1.5 anhand von konkreten Beispielen, jeweils unterschiedliche Erscheinungsformen für das zentrale Konzept der Kontraktionsbedingung vorgestellt und damit die Betrachtung von IFS motiviert.
- Mit dem gewonnenen Rüstzeug, werden in Kapitel 2 unterschiedliche Glättungsfunktionen untersucht und (minimale) Bedingungen zur Sicherung der gewünschten Ergodizitätseigenschaften herausgearbeitet (vgl. 2.1.1). Es werden die Beziehung bzw. Wechselwirkung der Glättungsfunktionen zu Kontraktionsbedingungen (2.1), aber auch zu Driftbedingungen und somit geometrischer Ergodizität (2.2) untersucht. Dieser Zusammenhang bietet die Möglichkeit, die unterschiedlichen Glättungskonzepte untereinander vergleichbar zu machen und so werden in Abschnitt 2.3 positive Effekte im Sinne von Parameterbereicherungen für geometrische Ergodizität betrachtet. Abschließend wird ein starker Bezug zu anderen Anwendungsgebieten herausgestellt, da sich das gleiche Instrumentarium dafür eignet und so einen Anknüpfungspunkt zu vorangegangenen Ergebnissen ergibt.
- In Kapitel 3 wird das empirische Fundament für die Nutzung von Glättungsfunktionen bereitet. So werden zunächst die Simulationsalgorithmen (3.1 - 3.3) für die unterschiedlichen Glättungsansätze erörtert und aufbauend darauf, die Parameterschätzung (3.4) in den unterschiedlichen Konstellationen des Glättens durchgeführt. Den Abschluss dieser computerunterstützten Untersuchung, bilden dann die Beurteilung der Parameterschätzungen und die Fusion der Teilergebnisse in einer Predict-Funktion.
- Auch wenn Kapitel 4 auf den ersten Blick vielleicht den Eindruck einer Sammlung von weniger wichtigen Randbemerkungen erwecken dürfte, vervollständigt es die Arbeit in ganz zentralen Punkten. Die ausführlichen Erläuterungen (z.B. 4.3) oder grundlegende bzw. ergänzende Beispiele (z.B. 4.5) wurden nur ausgelagert, um den Lesefluss nicht zu sehr zu stören und den ersten Blick auf das Gesamtgebilde nicht unnötig zu verdecken. Rückblickend bzw. fürs tiefere Verständnis sind sie jedoch von zentraler Bedeutung, was beim ersten Lesen durch die vielen Verweise auf den Anhang auch klar werden dürfte. Ein Einblick und gleichermassen Ausblick in die veränderte Ausgangssituation eines streng stationären Innovationsprozesses, bildet den Abschluss der Arbeit.

# 1 Einführung

## 1.1 Grundlagen

Die nachfolgenden Beispiele, sowie auch der Rest der Arbeit, stützen sich auf die Betrachtung von **IFS** bzw. **SRE**. Die Motivation für diese Vorgehensweise soll insbesondere das erste Kapitel verdeutlichen. Zum einen lassen sich unterschiedlichste Modelle oder Kriterien unter einen Hut bringen, wenn man sich von der speziellen Modell-Definition löst und diese zunächst einmal als Lösung eines IFS oder SRE sieht. Alle Ergebnisse auf dieser Ebene sind somit von genereller Natur und nicht an Details geknüpft. Weiter bietet sich auch die Möglichkeit, unterschiedlichste Herangehensweisen miteinander in Verbindung zu bringen, wenn man sie auf ihre Kernpunkte hin vergleicht, während sich der konkrete Weg dann doch sehr stark unterscheiden kann. In einigen Fällen wird der Umweg über den Nachweis von geometrischer Ergodizität gewählt, um die Existenz und Eindeutigkeit der stationären Verteilung zu sichern, während sich dies in anderen Fällen entweder erübrigt oder nicht offen zeigt. Dies führt dazu, dass man die Beziehung zwischen diesen Herangehensweisen nicht unbedingt auf den ersten Blick vermuten würde. Erst über den SRE/IFS-Ansatz lassen sich die starken Beziehungen zueinander offenlegen und die folgenden Ergebnisse inhaltlich sinnvoll einklammern.

### Definition 1.1.

Ein iteriertes Funktionensystem (**IFS** - iterated function system) wird beschrieben durch ein Tupel  $\{(W, \mathcal{B}_W), (X, \mathcal{X}), u, p\}$ , wobei

- (1)  $W$  ein metrischer Raum mit Metrik  $d$ , Borel- $\sigma$ -Algebra  $\mathcal{B}_W$  und  $(X, \mathcal{X})$  ein Messraum. Normalerweise bezeichnet man  $W$  als Zustandsraum und  $X$  als Ereignisraum.
- (2)  $(X_t, t \in \mathbb{N})$  ist ein streng stationärer Prozess mit  $X$ -wertigen Zufallsvariablen  $X_t$ , auch Innovationsprozess genannt,  $p$  ist die Randverteilung von  $X_t$ .
- (2') Oft betrachtet man  $(X_t, t \in \mathbb{N})$  als i.i.d. Folge von Zufallsvariablen, entsprechend repräsentiert durch ihre Verteilung  $p$ .
- (3)  $u : (W \times X, \mathcal{B}_W \otimes \mathcal{X}) \rightarrow (W, \mathcal{B}_W)$  ist eine messbare Abbildung und wird auch Übergangsfunktion des IFS genannt.

Bemerkenswert ist außerdem, dass für jedes  $x \in X$  durch  $u_x(\cdot) := u(\cdot, x)$  eine  $\mathcal{B}_W$ -messbare Abbildung  $u_x : W \rightarrow W$  entsteht. Bei gegebener Verteilung  $\mu$  einer Startvariablen  $W_0$ , wird ein stochastischer Prozess  $(W_t, t \in \mathbb{N}_0)$  generiert durch:

$$W_{t+1} = u(W_t, X_{t+1}), \quad t \in \mathbb{N}_0. \quad (\text{I})$$

Hierbei kann (I) auch als stochastische Rekursionsgleichung (**SRE**) aufgefasst werden, wobei in diesem Zusammenhang normalerweise  $t \in \mathbb{Z}$  gilt. Eine umfangreiche Quelle zu IFS, bildet die Arbeit von Iosifescu [16] und für weitere Informationen zu SRE sei verwiesen auf Straumann [28] oder Basrak [2].

Die Ausgangssituation unterscheidet sich in Abhängigkeit davon, ob nun  $t \in \mathbb{N}$  oder  $t \in \mathbb{Z}$  bzw. (2) oder (2') vorliegt, aber die Kernfrage nach Existenz und Eindeutigkeit einer streng stationären Lösung, sprich eines stochastischen Prozesses  $(W_t, t \in \mathbb{N} \text{ bzw. } t \in \mathbb{Z})$ , der (I) genügt, wird üblicherweise behandelt unter Zuhilfenahme eines Ergebnisses von Letac [21]. Genauer gesagt, erweist sich eine Kontraktionsbedingung als hinreichend für die Existenz und Eindeutigkeit einer stationären Verteilung für eine Markov-Kette.

**Satz 1.1.** (*Lemma von Letac - Principle in [21], Lemma in [16]*)

**Voraussetzungen:**

- (i) Die Abbildung  $u_x : W \rightarrow W$  ist stetig für alle  $x \in X$ .
- (ii)  $W_0^{w_0} = w_0$  und  $W_n^{w_0} = u(\dots(u(u(w_0, X_1), X_2), \dots), X_n)$  für  $n \in \mathbb{N}$ , ist eine Darstellung für (I).
- (iii)  $\widetilde{W}_0^{w_0} = w_0$  und  $\widetilde{W}_n^{w_0} = u(\dots(u(u(w_0, X_n), X_{n-1}), \dots), X_1)$  für  $n \in \mathbb{N}$ , ist ein sogenannter „backwards“-Ansatz und möglicherweise keine Markov-Kette, selbst wenn (ii) eine ist.  
Es gilt jedoch:  $W_n^{w_0} \sim \widetilde{W}_n^{w_0}$ ,  $w_0 \in W$ ,  $n \in \mathbb{N}_0$  (d.h.  $W_n^{w_0}$  und  $\widetilde{W}_n^{w_0}$  sind identisch verteilt).
- (iv)  $X_t, t \in \mathbb{N}$  seien unabhängig identisch verteilt.

**Aussage:**

Gilt  $\widetilde{W}_n^{w_0} \xrightarrow[n \rightarrow \infty]{f.s.} W_\infty$  und zwar unabhängig vom Startwert  $w_0$ , so ist  $\mu = \mathcal{L}(W_\infty)$  (Verteilung von  $W_\infty$ ) auch die einzige stationäre Verteilung von  $(W_n^{w_0})_{n \in \mathbb{N}}$ .

Im Kern der Sache ist Konvergenz in Verteilung  $\mathcal{L}(W_n^{w_0}) \xrightarrow[n \rightarrow \infty]{} \mathcal{L}(W_\infty)$  für den Ansatz (ii) ausreichend (vgl. Proposition [21]). Aus Gründen der mathematischen Handhabbarkeit nutzt man in der Regel aber die fast sichere Konvergenz des „backwards“-Ansatzes  $\widetilde{W}_0^{w_0}$  und erhält zusammen mit dem Kommentar in (iii) wiederum das Notwendige für (ii). Dieser verworrene Weg verbirgt sich letztendlich hinter dem überschaubaren Kriterium.

**Kriterium:** (vgl. [28], Variationen siehe [21])

Hinreichend für die fast sichere Konvergenz von  $\widetilde{W}_n^{w_0}$ :

$$u_t = u(\cdot, X_t) \text{ ist „Kontraktiv im Mittel“ („contractive on average“)}$$

in dem Sinne, dass

$$|u_t(w) - u_t(\hat{w})| \leq \Lambda(u_t)|w - \hat{w}| \quad \text{und} \quad \mathbb{E}[\log \Lambda(u_0)] < 0 \quad (\text{C1})$$

oder die abgeschwächte Variante

$$\mathbb{E}[\log \Lambda(u_0 \circ \dots \circ u_{-r+1})] < 0 \quad (\text{CR})$$

erfüllt ist.

Der dabei auftauchende Lipschitz-Koeffizient ist definiert durch:

$$\Lambda(u_t) = \sup_{\substack{w, \hat{w} \in W \\ w \neq \hat{w}}} \frac{|u_t(w) - u_t(\hat{w})|}{|w - \hat{w}|}$$

Eine Abbildung  $\phi : W \rightarrow W$  bezeichnet man als Lipschitz, wenn  $\Lambda(\phi)$  endlich ist und als Kontraktion, wenn  $\Lambda(\phi) < 1$  gilt. Der Index 0 in  $u_0$  ist hierbei zu verstehen als eine Art Platzhalter für die möglichen Auskommen der Übergangsfunktion und deutet an, dass die Abbildungen  $u_t$  typischerweise einen Lipschitzwert  $< 1$  hervorbringen müssen. Dieses Konstrukt ist also verantwortlich für die Namensgebung „Kontraktion im Mittel“, es wird nämlich keineswegs von ausnahmslos allen  $u_t$  die Kontraktionseigenschaft verlangt.

Die abgeschwächte Bedingung (CR) wird zum Beispiel in Iosifescu [16] und bei Straumann [28] ausgenutzt und bedeutet, dass bereits Kontraktivität nach  $r$  Iterationen ausreicht, um die fast sichere Konvergenz von  $\widetilde{W}_n^{w_0}$ , unabhängig vom Startwert  $w_0$ , zu garantieren. Die Abschwächung erahnt man mit der Submultiplikativität von  $\Lambda$ :  $(\Lambda(\phi \circ \psi) \leq \Lambda(\phi)\Lambda(\psi))$  für 2 Abbildungen  $\phi, \psi$ .

Es gibt durchaus Fälle, in denen  $\mathbb{E}[\log \Lambda(u_0)] \geq 0$  gilt, aber  $\mathbb{E}[\log \Lambda(u_0 \circ \dots \circ u_{-r+1})] < 0$  erfüllt ist, für ein  $r \in \mathbb{N}$ . Derartige Beispiele untermauern nochmals anschaulich den Aspekt des „hinreichenden Kriteriums“ und der Abschwächung der Bedingung - vergleiche dazu die eigenen Beispiele im Anhang (Abschnitt 4.1), sowie Straumann (3.3.1 - [28]).

Die zunächst befremdlich wirkende Vorgehensweise über den „backwards“-Ansatz, lässt sich an Hand eines Minimalbeispiels von Diaconis und Freedman [7] verdeutlichen. Die Arbeit von Diaconis und Freedman [7] wird darüber hinaus auch als zentrale Quelle für Straumann [28] hervorgehoben.

**Definition 1.2.** *Minimalbeispiel (6.1. - [7])*

*Betrachtet man  $0 < a < 1$ ,  $X_t$  i.i.d.,  $W_0 = w_0 \in \mathbb{R}$  und folgende Definitionsvorschrift eines autoregressiven Prozesses*

$$W_{t+1} = u(W_t, X_{t+1}) = aW_t + X_{t+1}, \quad t \in \mathbb{N}_0.$$

*Für den Ansatz (ii) bedeutet dies*

$$W_1^{w_0} = aw_0 + X_1, \quad W_2^{w_0} = a^2w_0 + aX_1 + X_2, \quad W_3^{w_0} = a^3w_0 + a^2X_1 + aX_2 + X_3, \quad W_4^{w_0} = \dots$$

*und für (iii) wiederum*

$$\widetilde{W}_1^{w_0} = aw_0 + X_1, \quad \widetilde{W}_2^{w_0} = a^2w_0 + X_1 + aX_2, \quad \widetilde{W}_3^{w_0} = a^3w_0 + X_1 + aX_2 + a^2X_3, \quad \widetilde{W}_4^{w_0} = \dots$$

*Der Unterschied wird bei einer Betrachtung über die Zeit sehr deutlich. Während im ersten Fall („forward“-Ansatz) zum Zeitpunkt  $n$  der Zufall durch  $X_n$  voll zum tragen kommt, wird er im zweiten Fall („backwards“-Ansatz) durch den Vorfaktor  $a^n$  abgeschwächt. Aus derartigen Effekten entsteht das unterschiedliche Konvergenzverhalten und macht insbesondere bei autoregressiven Prozessen den „backwards“-Ansatz zum nützlichen Hilfsmittel.*

Im Folgenden soll anhand einiger Beispiele, die jeweilige Umsetzung der obigen Bedingungen genauer untersucht werden. Die unterschiedlichen Hauptbedingungen würden auf den ersten Blick vielleicht gar keinen Zusammenhang zwischen den Beispielen vermuten lassen. Erst wenn man diese einander gegenüberstellt und immer die Beziehung zum Lemma von Letac (Satz 1.1) im Hinterkopf hat, erkennt man ihre starke Beziehung zueinander.

## 1.2 GARCH - Beispiel

Das bisher Gesehene spiegelt sich beispielsweise in den Ergebnissen von Bougerol und Picard [25] zur Stationarität von GARCH-Prozessen wider und veranschaulicht eine erste Umsetzungs- bzw. Anwendungsmöglichkeit der obigen Kontraktionsbedingung.

Betrachtet man eine i.i.d. Folge zufälliger Matrizen  $((A_t, B_t))_{t \in \mathbb{Z}}$ , wobei  $A_t$   $d \times d$ -Matrizen und  $B_t$   $d$ -dimensionale Vektoren sind, so bildet

$$Y_t = A_t Y_{t-1} + B_t, \quad t \in \mathbb{Z} \quad (1)$$

den Modell-Rahmen um GARCH(p,q)-Prozesse zu behandeln und gleichermaßen nur eine besondere Darstellung für eine SRE ( $Y_t = u_t(Y_{t-1})$ ;  $u_t(y) = A_t y + B_t$ ). Im Anhang (Abschnitt 4.1.2) wird diese Konstruktion im Detail nachvollzogen, da der GARCH(p,q)-Prozess als Beispiel für die abgeschwächte Bedingung nach  $r$  Iterationen (CR) herangezogen werden kann. An dieser Stelle interessiert nur das Kriterium bzw. das resultierende Ergebnis und somit reicht eine grobe Vorstellung von den Matrizen, ohne sich im Klaren sein zu müssen wie ein GARCH-Prozess im Detail konstruiert ist und wie sich dies im Aufbau der Matrizen niederschlägt:

$$A_t = \begin{pmatrix} \alpha_1 Z_t^2 + \beta_1 & \beta_2 & \cdots & \beta_{q-1} & \beta_q & \alpha_2 & \alpha_3 & \cdots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ Z_t^2 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad t \in \mathbb{Z}.$$

Die Definitionsvorschrift eines GARCH(p,q)-Prozesses (siehe Anhang 4.1.2) wird in die Gestalt einer stochastischen Matrix übersetzt und somit die Möglichkeit geschaffen, die Suche nach einer stationären Verteilung mit Hilfe eines völlig anderem Instrumentariums anzugehen. Mit dem Produkt von stochastischen Matrizen kann man zu Ergebnissen gelangen, die man an Hand der eigentlichen Gestalt des Prozesses nicht hätte ablesen können.

Man spricht davon, dass ein  $d$ -dimensionaler stochastischer Prozess  $(Y_t)$  einer **stochastischen Rekursionsgleichung (SRE)** genügt oder Lösung einer **SRE (stochastic recurrence equation)** ist. Um die Existenz und Eindeutigkeit einer stationären Lösung zu sichern, wird in [25] die Eigenschaft „contractive on average“ gezeigt und zwar indem die Negativität des Lyapunov-Exponenten gesichert wird:

$$\gamma = \inf_{n \in \mathbb{N}_0} \left\{ \frac{1}{n+1} \mathbb{E}(\log \|A_0 \cdots A_{-n}\|_{op}) \right\} < 0.$$

Hierbei bezeichnet  $\|\cdot\|_{op}$  die zur euklidischen Norm zugehörige Operatornorm, wobei die Wahl der Norm keinen Einfluss hat auf den Wert von  $\gamma$ .

**Satz 1.2. (Theorem 1.3 - [25])**

*Der GARCH(p,q)-Prozess besitzt genau dann eine streng stationäre Lösung, wenn der Lyapunov-Exponent negativ ist. Diese Lösung ist darüber hinaus ergodisch, d.h. neben dem konstanten Verhalten über dem Zeitstrahl, können auch statistisch relevante Werte, wie beispielsweise der Erwartungswert, aus einer einzigen hinreichend langen Realisation gewonnen werden. Sofern der Störterm vorgegeben ist, so ist dies auch die einzige stationäre Lösung.*

Das Ergebnis ist für sich genommen bereits äußerst interessant, da es sowohl hinreichend als auch notwendig ist, doch der entscheidende Faktor für die hier angestrebte Betrachtung ist der enge Zusammenhang zur Kontraktionsbedingung.

Zusammen mit

$$\|A\|_{op} = \sup_{x \neq 0} \frac{|Ax|}{|x|}; \quad \Lambda(u) = \sup_{x \neq y} \frac{d(u(x), u(y))}{d(x, y)} \quad (\text{Metrik induziert durch die euklidische Norm})$$

folgt insbesondere

$$\Lambda(u_0 \circ \cdots \circ u_{-r+1}) = \|A_0 \cdots A_{-r+1}\|_{op}$$

und damit aus der Forderung an den Lyapunov-Exponenten auch die abgeschwächte Bedingung für das Letac Lemma (Satz 1.1). Hinter der Nutzung des Lyapunov-Exponenten (Hauptbedingung  $\gamma < 0$ ) verbirgt sich auf den zweiten Blick also nichts anderes, als die obigen Ausführungen zu Kontraktionsbedingungen. Der vermeintliche Unterschied im verwendeten Hauptkriterium ist auf den Kunstgriff zurückzuführen, dass man den GARCH(p,q)-Prozess mittels einer affinen Übertragungsfunktion formuliert. Mit dieser Vorgehensweise kommt man zu einem notwendigen und hinreichenden Kriterium, der Bezug zur Kontraktionsbedingung erschliesst sich dadurch jedoch erst auf den zweiten Blick.

In Straumann [28] - 3.3.1 und Basrak [2] - 3.2.1. werden unter anderem auch AGARCH(p,q)-Modelle in die Betrachtung eingebunden bzw. die Ergebnisse darauf erweitert. Hierbei bezieht sich das A auf Asymmetrie und damit das unterschiedliche Verhalten bei positiven oder negativen Veränderungen (Shocks), ein bei Finanzmarktdaten oft zu beobachtender Umstand.

Die konkrete Berechnung des Lyapunov-Exponenten kann im allgemeinen sehr schwierig werden, wie bereits „einfache“ bzw. seit Jahrzehnten bekannte und in der Literatur verbreitete Modelle zeigen (vgl. z.B. GARCH(p,q), generalized AR(p) - [28], [25]). Mit Hilfe von Gleichung (2) und Monte-Carlo-Simulationen kann man in vielen Fällen jedoch zu einer guten Näherung kommen.

Nimmt man neben der Negativität des Lyapunov-Exponenten weiter an  $\mathbb{E} [\log^+ \|A_0\|_{op}] < \infty$ , so gilt:

$$\gamma = \lim_{n \rightarrow \infty} \frac{1}{n+1} \log \|A_0 \cdots A_n\|_{op} \text{ f.s.} \quad (2)$$



Einen Simulationsansatz oder eine komplizierte Berechnung kann man in so mancher Situation aber auch umgehen. So ist  $\gamma < 0$  bereits erfüllt, wenn  $\mathbb{E}(\log \|A_0\|_{op}) < 0$  gilt (betrachte die Definition von  $\gamma$  mit  $n = 0$ ). Für die Dimension  $d=1$  folgt weiter  $\gamma = \mathbb{E} \log |A_0|$ .

Im Falle einer deterministischen Matrix  $A$ , verringert sich der Rechenaufwand auf:

$$\gamma = \log \rho(A),$$

wobei  $\rho(A)$  den Spektralradius der Matrix  $A$  bezeichnet. Dies ergibt sich insbesondere aus der Darstellung des Spektralradius

$$\rho(A) = \lim_{n \rightarrow \infty} \|A^n\|^{1/n}$$

für irgendeine Matrix-Norm. Zusammenfassend erhält man:

$$A \text{ deterministisch} \Rightarrow [\gamma < 0 \Leftrightarrow \rho(A) < 1].$$

In Proposition 3.3.3. - Straumann [28] finden sich noch weitere Beispiele, die in ihrer Gesamtheit eindrucksvoll verdeutlichen welche Anstrengungen man unter Umständen unternehmen muss, um ein zunächst unscheinbares Kriterium zu erfüllen.

### 1.3 Basrak - Beispiel

In Anbetracht der aufwendigen Berechnung des Lyapunov-Exponenten, bietet Basrak [2] eine für bestimmte Situationen vielleicht interessante Alternative und somit eine weitere Erscheinungsform für die gleiche Grundidee. Für Prop. 3.2.9 - [2] wird unter anderem  $h_A(\varepsilon) < 1$  gefordert, um die stärkere Konvergenz geometrische Ergodizität zu erreichen. Aus geometrischer Ergodizität kann die Existenz sowie Eindeutigkeit einer stationären Anfangsverteilung gefolgert werden und so auch die stationäre Lösung der entsprechenden SRE.

Mit geometrischer Ergodizität gewinnt man ein Informationsplus über den betrachteten stochastischen Prozess, da eine Aussage über die Konvergenzgeschwindigkeit gegen die stationäre Verteilung möglich ist. Konvergenz bedeutet hierbei bezüglich der t.v.n. (total variation norm).

**Definition 1.3.** *total variation norm - t.v.n.*

Für ein Maß  $\mu$  auf  $\mathcal{B}_W$  ist die total variation norm  $\|\cdot\|$  definiert durch:

$$\|\mu\| := \sup_{f:|f|\leq 1} |\mu(f)| = \sup_{f:|f|\leq 1} \left| \int f(x)\mu(dx) \right| = \sup_{A \in \mathcal{B}_W} \mu(A) - \inf_{A \in \mathcal{B}_W} \mu(A)$$

Betrachtet man zwei Wahrscheinlichkeitsmaße  $\nu_1$  und  $\nu_2$ , so ergibt sich

$$\|\nu_1(\cdot) - \nu_2(\cdot)\| = 2 \sup_{A \in \mathcal{B}_W} |\nu_1(A) - \nu_2(A)|$$

und dieser Fall wird auch als total Variation distance bezeichnet.

**Definition 1.4.** *Ergodizität*

Eine Markov-Kette  $\{W_t\}$  heißt *ergodisch*, wenn ein invariantes Wahrscheinlichkeitsmaß  $\pi$  existiert, so dass

$$\lim_{n \rightarrow \infty} \|P^n(w, \cdot) - \pi(\cdot)\| = 0 \quad \forall w \in W.$$

Ergodizität bedeutet demnach, dass die Übergangswahrscheinlichkeiten  $P^n(w, \cdot)$  gegen eine stationäre Verteilung  $\pi$  konvergieren (bzgl. t.v.n.). Die Konvergenz geschieht hierbei unabhängig vom Startwert  $W_0 = w$  und bedeutet ein stabiles Verhalten der Markov-Kette für  $n$  hinreichend groß. Eine wichtige wenn auch offensichtliche Folgerung ist die Eindeutigkeit des stationären Maßes  $\pi$ .

Angenommen  $\pi^*$  wäre ein weiteres Maß, so folgt

$$\begin{aligned} \|\pi(\cdot) - \pi^*(\cdot)\| &= \|\pi(\cdot) - P^n(w, \cdot) + P^n(w, \cdot) - \pi^*(\cdot)\| \\ &\leq \|P^n(w, \cdot) - \pi(\cdot)\| + \|P^n(w, \cdot) - \pi^*(\cdot)\| \rightarrow 0 \end{aligned}$$

und damit  $\pi = \pi^*$ .

Während die Ergebnisse von Bougerol und Picard [25] genau diesen Bedingungen genügen und man so die Konvergenz (bzgl. t.v.n.) des stochastischen Prozesses gegen eine stationäre Verteilung gesichert hat, kann man keine Auskunft über das Tempo dieses Konvergenzverhaltens machen. Dazu bedarf es einer Erweiterung des Ergodizitätsbegriffs oder mit anderen Worten eines stärkeren Konzepts.

**Definition 1.5.** *geometrische Ergodizität*

Eine Markov-Kette  $\{W_t\}$  heißt *geometrisch ergodisch*, wenn folgendes gilt:  $\exists$  ein invariantes Wahrscheinlichkeitsmaß  $\pi$ ,  $\gamma \in (0, 1)$  und eine  $\pi$ -integrierbare nicht-negative Funktion  $h$ , so dass

$$\|P^n(w, \cdot) - \pi(\cdot)\| \leq h(w)\gamma^n \quad \forall w \in W.$$

Weitere Ergodizitätsbegriffe werden beispielsweise in [34] behandelt.

In diesem Kontext lassen sich die Ergebnisse nun einbetten und in Verbindung setzen zu dem Vorangegangenen. Für eine Zufallsmatrix  $A$ , definiert man

$$h_A(v) = \mathbb{E}\|A\|^v \tag{3}$$

für  $v \in [0, \infty)$ . Unter der Bedingung  $\mathbb{E}\|A\|^\nu < \infty$  für ein  $\nu > 0$ , bildet  $h_A$  eine wohldefinierte reellwertige Funktion in dem Intervall  $[0, \nu]$ . Die ersten beiden Ableitungen ergeben sich zu

$$h'_A(v) = \mathbb{E}\|A\|^v \ln \|A\|$$

und

$$h''_A(v) = \mathbb{E}\|A\|^v \ln^2 \|A\|,$$

was  $h_A$  als konvexe Funktion über  $[0, \nu]$  auszeichnet ( $h_A''(v) > 0$ ).

Die Anknüpfung zu oben ergibt sich durch die Aussage von Lemma 3.2.7. - [2].  
Demnach ist die Bedingung

$$h_A(\varepsilon) < 1 \quad (4)$$

erfüllt für ein  $\varepsilon > 0$ , genau dann, wenn

$$\mathbb{E} \ln \|A\| < 0 \quad \text{und} \quad h_A(\delta) < \infty$$

für ein  $\delta > 0$  gilt. Insbesondere wäre unter diesen Bedingungen der Lyapunov-Exponent  $< 0$  und damit auch das Lemma von Letac (Satz 1.1) erfüllt.

Je nach Ausgangssituation kann die Berechnung von (3) oder die Überprüfung von (4) wesentlich unkomplizierter sein, als die Beschäftigung mit dem Lyapunov-Exponenten oder der Rückzug auf  $r$ -fache Iterationen. Mit dem Kriterium von Basrak hat man es im Kern der Sache aber trotzdem mit einem Kontraktionskriterium zu tun.

## 1.4 SLARM - Beispiel

Weitere interessante Abwandlungen/Erscheinungsformen findet man in der Arbeit von Rudolph [26] - SLARM (stochastic linear autoregressive models), in der die euklidische Norm und die zugehörige Operatornorm weiter in den Fokus gerückt werden.

Ein SLARM wird definiert durch die Funktionsgleichung

$$W_t = F_t W_{t-1} + G U_t$$

$((F_t, U_t), t \in \mathbb{N}, \mathbb{Z})$  i.i.d. mit einer üblichen Verteilung  $\mu$ .

Die Parallelen zur SRE in (1) sind unverkennbar und auch die Umsetzung des GARCH(p,q)-Falls in **Example 2** - [26] unterstreicht diesen starken Zusammenhang - letztendlich wird die Rekursionsidee auf spezielle, da lineare, Prozesse eingeschränkt. Wichtige Unterschiede zu dem Bisherigen ergeben sich wiederum beim Hauptkriterium, aber auch in Hinblick auf die Abstufung der Ergebnisse. So sind die erreichbaren Schlussfolgerungen im hohen Maße davon abhängig, welche Zusatzbedingungen gemeinsam mit dem Hauptkriterium gefordert werden.

Betrachtet man  $u_t(x) = F_t(x) + G U_t$  für  $t \in \mathbb{N}, \mathbb{Z}$ , so erhält man für ein SLARM die Darstellung

$$W_t = u_t(W_{t-1}), \quad t \in \mathbb{N}, \mathbb{Z}$$

und für  $u : \mathbb{R}^k \rightarrow \mathbb{R}^k$  mit  $u(x) = Fx + GU$  folgt damit

$$\Lambda(u) = \sup_{x \neq y} \frac{\|F(x - y)\|}{\|x - y\|} = \sup_{v \neq 0} \frac{\|Fv\|}{\|v\|} = \|F\|_{op}.$$

Das ergibt als Hauptbedingung in [26]

$$\mathbb{E}(\|F_0\|_{op}) < 1 \quad (\text{SEVF})$$

und so wiederum auch  $\mathbb{E}(\log \Lambda(u_0)) < 0$  (Lemma von Letac - Satz 1.1), denn nach der Jensen'schen Ungleichung gilt:

$$\mathbb{E}(\log \Lambda(u_0)) \leq \log \mathbb{E}(\Lambda(u_0)) = \log \mathbb{E}(\|F_0\|_{op}).$$

Für den Fall  $F_t = F$  konstant, kann man mit

$$\rho(F) \leq \|F\|_{op}$$

schlussfolgern, dass (SEVF) auch  $\rho(F) < 1$  nach sich zieht. Vgl. dazu auch die Anmerkung im Zusammenhang mit dem Lyapunov Exponenten, denn (SEVF)  $\Rightarrow \gamma < 0$ ,  $\gamma = \log \rho(A)$ .

Für  $(W_t, t \in \mathbb{Z})$  ergibt sich ein streng stationärer Prozess, also ein „stabiles“ Verhalten über den gesamten Zeitstrahl. Mit einer endlichen Vergangenheit  $(W_t, t \in \mathbb{N}_0)$ ,  $((F_t, U_t), t \in \mathbb{N})$  stellt sich die Situation zwar anders da, aber mit

$$l := \sup_{x \neq y} \int \frac{\|F(x - y)\|}{\|x - y\|} \mu(d(F, U))$$

und Iosifescu [16] - S.7

$$l \leq \int \Lambda(u) \mu(d(F, U)) = \mathbb{E}(\|F\|_{op})$$

kann (SEVF) auch als Kriterium für die Konvergenz des stochastischen Prozesses verwendet werden. In dieser Konstellation sichert man Ergodizität und damit findet eine Stabilisierung im Sinne von Konvergenz gegen eine stationäre Verteilung statt - vgl. oben. Festzuhalten bleibt, dass mit dem selben Kriterium einmal die Stationarität der Lösung und das andere Mal eine Stationarisierung gesichert wird, je nachdem ob man  $t \in \mathbb{Z}$  oder  $\mathbb{N}$  betrachtet.

Mit der euklidische Norm  $\|x\| = \sqrt{\sum_{i=1}^k x_i^2}$  und deren besonderen Aufbau, erhält man folgende Abschätzung

$$l = \sup_{x \neq y} \frac{\int \sqrt{(x - y)' F' F (x - y)} \mu(d(F, U))}{\sqrt{(x - y)' (x - y)}} \leq \sup_{x \neq y} \frac{\sqrt{(x - y)' \int (F' F \mu(d(F, U))) (x - y)}}{\sqrt{(x - y)' (x - y)}} \\ \left( = \sup_{x \neq y} \frac{\|\mathbb{E}(F' F)(x - y)\|}{\|x - y\|} = \|\mathbb{E}(F' F)\|_2 = \sqrt{\lambda_{max}(\mathbb{E}(F' F))} = \sqrt{\rho(\mathbb{E}(F' F))} \right)$$

und damit

$$l \leq \sqrt{\rho(\mathbb{E}(F' F))}.$$

Für ein SLARM

$$W_t = F_t W_{t-1} + G U_t, \quad t \in \mathbb{N}$$

mit dem Startvektor  $W_0$ , ergibt sich so wiederum eine Abwandlung der Hauptbedingung:

$$\rho(\mathbb{E}(F_1' F_1)) < 1 \quad (\text{SEVF}')$$

Diese Hauptbedingungen führen mit ein paar weiteren kleinen Forderungen, die beispielsweise die Aperiodizität und Irreduzibilität der Markov-Kette sichern, zu geometrischer Ergodizität (Theorem 3/4 in [26]). Man vergleiche dazu insbesondere mit Basrak [2] 2.8 - S. 79 und 3.2.9 - S. 101.

Die bisherigen Beispiele zeigen, wie sich eine Umsetzung auf ein konkretes Problem darstellen und insbesondere im Detail auch unterscheiden kann. So sind trotz sehr ähnlicher oder identischer Hauptbedingung, je nach Begleitumständen, unterschiedliche Schlussfolgerungen möglich. Eine klare Abstufung, von der stärksten bis zur schwächsten Bedingung, ist nicht immer zweifelsfrei erkennbar, der zentrale Kern ist jedoch immer der Gleiche und so soll das folgende Beispiel die Betrachtung dahingehend abschliessen, dass möglichst allgemeingültige Ergebnisse, unabhängig vom konkreten untersuchten Modell festgehalten werden.

## 1.5 Iosifescu - Beispiel

Die oben verwendeten Bedingungen zu IFS (i.i.d. case), werden in Iosifescu [16] noch weiter verallgemeinert und abgeschwächt. Unter der stillschweigend vorausgesetzten Stetigkeitsannahme von  $u_x : W \rightarrow W$  für  $p$ -fast alle  $x \in X$ , definiert man zunächst

$$l(x) = l(x, d) = s(u_x) := \sup_{\substack{w' \neq w'' \\ w', w'' \in W}} \frac{d(u_x(w'), u_x(w''))}{d(w', w'')}, \quad x \in X.$$

Ist der metrische Raum  $W$  auch separabel, dann ist die Abbildung  $x \rightarrow l(x)$  von  $X$  nach  $\overline{\mathbb{R}}$   $(\mathcal{X}, \mathcal{B}_{\overline{\mathbb{R}}})$ -messbar. Weiter fordert man

$$l := \sup_{\substack{w' \neq w'' \\ w', w'' \in W}} \int_X \frac{d(u_x(w'), u_x(w''))}{d(w', w'')} p(dx) < 1, \quad (2.4)$$

$(l \leq \int_X l(x) p(dx))$  und für ein  $w_0 \in W$

$$\int_X d(w_0, u_x(w_0)) p(dx) < \infty. \quad (2.5)$$

Die Bedingungen (2.4)<sup>1</sup>, (2.5) sichern die Existenz und Eindeutigkeit einer stationären Lösung (vgl. Iosifescu [16], Theorem 3.3).

Für  $0 < \alpha \leq 1$  ist auch  $d^\alpha$  wiederum eine Metrik und man erhält eine Verallgemeinerung der obigen Bedingungen durch:

$$l_\alpha := \sup_{\substack{w' \neq w'' \\ w', w'' \in W}} \int_X \frac{d^\alpha(u_x(w'), u_x(w''))}{d^\alpha(w', w'')} p(dx) < 1, \quad (3.6)$$

sowie

$$\int_X d^\alpha(w_0, u_x(w_0)) p(dx) < \infty \quad (3.7)$$

---

<sup>1</sup>Zur besseren Orientierung, entspricht die Nummerierung derjenigen in [16].

für  $w_0 \in W$ .

Die Bedingung der „Kontraktivität im Mittel“ (3.6) wird nun, wie zuvor durch Forderung der Kontraktivität nach  $n$  Schritten, abgeschwächt. Letztlich konstruiert man durch die Festlegung  $u_{x(n)}(w') = u(\dots u(u(w', x_1), x_2) \dots, x_n)$  ein neues IFS, wobei  $p_n$  das  $n$ -fache Produktmaß von  $p$  bezeichnet und die Bedingungen (3.6), (3.7) sich ergeben zu

$$l_{\alpha,n} := \sup_{\substack{w' \neq w'' \\ w', w'' \in W}} \int_X \frac{d^\alpha(u_{x(n)}(w'), u_{x(n)}(w''))}{d^\alpha(w', w'')} p_n(dx^{(n)}) < 1, \quad (3.15)$$

$$\int_{X^{n_0}} d^\alpha(w_0, u_{x(n_0)}(w_0)) p_{n_0}(dx^{(n_0)}) < \infty. \quad (3.16)$$

Die tatsächliche Abschwächung erkennt man wiederum mit der Submultiplikativität

$$l_{\alpha,m+n} \leq l_{\alpha,m} l_{\alpha,n}, \quad n, m \in \mathbb{N}.$$

Damit können analog Aussagen zur Stationarität für das neue IFS gemacht werden. Will man diese Schlüsse auf das Ausgangs-IFS und die damit verbundene Markov-Kette übertragen, so erfordert dies zusätzliche Annahmen (vgl. Corollary 3.7 - [16]).

Zusammenfassend stellen also die obigen Bedingungen:

- Bougerol und Picard [25]  $\leftrightarrow$  Lyapunov-Exponent.
- Basrak [2]  $\leftrightarrow h_A(\varepsilon) < 1$ .
- Rudolph [26] bzw. Iosifescu [16]

eine Liste von möglichen Konstellationen zur Erfüllung der Kontraktionsbedingung oder des Lemmas von Letac (Satz 1.1) dar. Die Hauptbedingungen sind an die jeweilige Problemsituation angepasst und unterscheiden sich auf den ersten Blick natürlich beträchtlich. Auf Grund der unterschiedlichen Begleitumstände, sind die Bedingungen nicht immer eindeutig als stärker/schwächer im Vergleich zueinander erkennbar. Mit der Zielsetzung die Existenz einer stationären Anfangsverteilung zu sichern, liegt ihnen aber immer der gleiche Kern zu Grunde. Weitere Variationen werden folgen.

Verändert man den Innovationsprozess  $(X_n, n \in \mathbb{N})$  nun dahingehend, dass man nicht mehr den i.i.d. Fall, sondern nur noch einen streng stationären Prozess betrachtet, so hat dies gravierende Unterschiede zur Folge.

Strenge Stationarität bedeutet hierbei, dass die Verteilung des Vektors

$$(\xi_m, \dots, \xi_{m+k})$$

$m \in \mathbb{N}_+, k \in \mathbb{N}$ , nicht von  $m$  abhängt. Typischerweise behandelt man streng stationäre Prozesse unter der zusätzlichen Annahme einer unendlichen Vergangenheit (vgl. Prop. 4.6 - [16]).

Anders als im i.i.d. Fall, hat man es nun nicht mehr mit einer Markov-Kette zu tun. Die fehlende Markov-Eigenschaft macht eine andere Vorgehensweise und ein vollkommen anderes Instrumentarium notwendig. Dieser inhaltlichen Trennung ist es geschuldet, dass zunächst nur der i.i.d Fall behandelt wird.

## 2 AR(1)-Modell mit ARCH(1)-Innovationen

Wir kommen nun zu einem, für weite Teile der vorliegenden Arbeit, zentralen Modell. Dieses zeichnet sich, trotz der auf den ersten Blick vielleicht unkompliziert wirkenden Konstruktion, durch eine hohe Variationsfähigkeit aus und soll als Ausgangspunkt für die Überlegungen zu Glättungsansätzen dienen. Im ersten Fall wird auf Grundlage der bisher besprochenen Bedingungen, geometrische Ergodizität gefolgert, was wiederum Existenz und Eindeutigkeit einer stationären Lösung (Anfangsverteilung) nach sich zieht. Das Ausgangsmodell und die Frage nach der geometrischen Ergodizität, bleiben auch im zweiten Fall erhalten, doch die Vorgehensweise wird in eine andere Richtung führen und den Grundstein für weitere Betrachtungen bilden.

Sowohl in Klüppelberg und Peng [24] als auch in Borkovec und Klüppelberg [19], wird ein AR(1)-Modell mit ARCH(1)-Fehler/Innovation betrachtet:

$$W_t = \beta W_{t-1} + \sqrt{\alpha_0 + \alpha_1 W_{t-1}^2} \varepsilon_t, \quad t \in \mathbb{N}$$

wobei  $\beta \in \mathbb{R}$ ,  $\alpha_0 > 0$ ,  $\alpha_1 \geq 0$ ,  $\{\varepsilon_t : t \in \mathbb{N}\}$  i.i.d. mit Erwartungswert 0 und Varianz 1. Weiter sei  $W_0$  unabhängig von  $\{\varepsilon_t : t \in \mathbb{N}\}$ . Das obige Modell bildet offenbar eine Markov-Kette, so dass geometrische Ergodizität von  $\{W_t\}$  die Existenz und Eindeutigkeit einer stationären Anfangsverteilung sichert. Die geforderten Bedingungen in Borkovec und Klüppelberg [19] (im folgenden oft mit BuK abgekürzt) lauten:

(B1) Der Störterm  $(\varepsilon_t, t \in \mathbb{N})$  besitze eine symmetrische, strikt positive und stetige Lebesgue-Dichte über  $(-\infty, +\infty)$ , sowie ein existentes zweites Moment  $\mathbb{E}(\varepsilon_t^2) < \infty$ .

(B2) Der Parameterraum sei

$$\Theta = \{\theta = (\beta, \alpha_0, \alpha_1)^T : \mathbb{E}(\log |\beta + \sqrt{\alpha_1} \varepsilon|) < 0, -\infty < \beta < +\infty, \alpha_0 > 0, \alpha_1 \geq 0\}.$$

BuK [19] leiten die Bedingung (B2) aus der Betrachtung einer Funktion  $h_{\beta, \alpha_1} : [0, \infty) \rightarrow [0, \infty)$  für  $\beta \in \mathbb{R}$  und  $\alpha_1 > 0$  ab (vgl. auch Proposition 2 - BuK [19]):

$$h_{\beta, \alpha_1}(v) := \mathbb{E}(|\beta + \sqrt{\alpha_1} \varepsilon|^v), \quad v \geq 0.$$

Die Funktion  $h_{\beta, \alpha_1}(\cdot)$  ist dabei streng konvex über dem Intervall  $[0, T)$ , wobei

$$T := \inf\{v \geq 0 \mid \mathbb{E}(|\sqrt{\alpha_1} \varepsilon|^v) = \infty\}.$$

Die Forderung an die Parameter  $\beta$  und  $\alpha_1$ , entstammt dabei der Bedingung

$$h'_{\beta, \alpha_1}(0) = \mathbb{E}(\log |\beta + \sqrt{\alpha_1} \varepsilon|) < 0.$$

Man vergleiche dieses Konstrukt insbesondere mit dem Ergebnis von Basrak [2] s.o. und beachte, dass im Allgemeinen eine genauere Angabe der Parameterpaarungen nicht möglich ist. Für den Fall  $\varepsilon \sim \mathcal{N}(0, 1)$  bieten BuK [19] auf den Seiten 1226/1227 jedoch einige tabellierte Werte an, auf die wir später zurückkommen werden.

Wertebereich von  $\alpha_1$  bei gegebenem  $|\beta|$  unter den Bedingungen  $h'_{\beta, \alpha_1}(0) < 0$  und  $\varepsilon \sim \mathcal{N}(0, 1)$  :

$ \beta $	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
$\alpha_1$	(0, 3.56]	(0, 3.55]	(0, 3.52]	(0, 3.47]	(0, 3.39]	(0, 3.30]	(0, 3.18]	(0, 3.04]
$ \beta $	0.8	0.9	1	1.1	1.2	1.25	1.27	1.27805
$\alpha_1$	(0, 2.87]	(0, 2.66]	(0, 2.42]	(0.17, 2.11]	(0.38, 1.69]	(0.58, 1.38]	(0.75, 1.19]	(0.94, 0.96]

Supremum von  $|\beta|$  bei gegebenem  $\alpha_1$  unter den Bedingungen  $h'_{\beta, \alpha_1}(0) < 0$  und  $\varepsilon \sim \mathcal{N}(0, 1)$  :

$\alpha_1$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$ \beta $	1.05	1.11	1.16	1.20	1.23	1.25	1.26	1.27	1.28
$\alpha_1$	1	1.1	1.2	1.5	2	2.5	3	3.5	3.56
$ \beta $	1.28	1.27	1.27	1.23	1.13	0.97	0.72	0.24	0.04

Was sofort ins Auge fällt, ist wohl der Umstand, dass zwar  $\alpha_0 > 0$  gefordert wird, aber in (B2) dann nicht weiter Beachtung findet. Diese verwirrende Forderung und auch die Gestalt der Bedingung (B2) wird einem klar, wenn man sich in Hinblick auf IFS das AR(1)-ARCH(1)-Modell als Abbildung:  $W \times X \rightarrow W$  vorstellt

$$r(w, x) = \beta w + \sqrt{\alpha_0 + \alpha_1 w^2} x$$

und sie anschließend für die zuvor schon mehrfach erwähnte Bedingung  $\mathbb{E}[\log \Lambda(r_0)] < 0$  auswertet. Der Lipschitz-Koeffizient ergibt sich demnach zu

$$\begin{aligned}
& \sup_{\substack{w, \hat{w} \neq 0 \\ w \neq \hat{w}}} \frac{|\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} x - \beta \hat{w} - \sqrt{\alpha_0 + \alpha_1 \hat{w}^2} x|}{|w - \hat{w}|} \\
&= \sup_{\substack{w, \hat{w} \neq 0 \\ w \neq \hat{w}}} \frac{|\beta(w - \hat{w}) + (\sqrt{\alpha_0 + \alpha_1 w^2} - \sqrt{\alpha_0 + \alpha_1 \hat{w}^2}) x|}{|w - \hat{w}|} \\
&\quad \text{O.B.d.A. sei } w^2 > \hat{w}^2, \quad \sqrt{\cdot} \text{ monoton wachsend; Abl. pos. und fallend} \\
&\leq \sup_{\substack{w, \hat{w} \neq 0 \\ w \neq \hat{w}}} \frac{|\beta(w - \hat{w}) + (\sqrt{\alpha_1 w^2} - \sqrt{\alpha_1 \hat{w}^2}) x|}{|w - \hat{w}|} \\
&= \sup_{\substack{w, \hat{w} \neq 0 \\ w \neq \hat{w}}} \frac{|\beta(w - \hat{w}) + \sqrt{\alpha_1} x (w - \hat{w})|}{|w - \hat{w}|} \\
&\leq \sup_{\substack{w, \hat{w} \neq 0 \\ w \neq \hat{w}}} \frac{|\beta + \sqrt{\alpha_1} x| |w - \hat{w}|}{|w - \hat{w}|} = |\beta + \sqrt{\alpha_1} x|
\end{aligned}$$

Die von BuK [19] gestellte Bedingung für geometrische Ergodizität, fügt sich also wunderbar in die Liste der Abwandlungen der Kontraktionsbedingung ein.

Im Wirtschaftsgeschehen oder ganz konkret am Beispiel des Aktienkurses eines etablierten Unternehmens wird jedem sofort klar sein, dass Schwankungen zwar immer gegeben sind, aber



insbesondere bei längerer Betrachtung ein gewisses Spektrum nicht verlassen wird. So erscheint es äußerst unwahrscheinlich, ja geradezu unmöglich, dass ein bestimmter „niedrigster“ Wert unterschritten bzw. ein „höchster“ Wert überschritten wird. In aller Regel werden derartige Ausreißer nur unter Katastrophenszenarien zu erwarten sein und sollten deshalb für die Modellierung des Alltagsgeschehens keine bzw. eine sehr geringe Bedeutung haben. Es erscheint nur sinnvoll, Schranken wie die oben beschriebenen gleich in die Modellbildung einfließen zu lassen, um sie so realitätsnah wie möglich/nötig zu machen.

Herkenrath [13] erfasst diesen Grundgedanken in der Modellvorstellung eines IFS

$$\{(W, \mathcal{B}_W), (X, \mathcal{X}), u, p = \mathcal{N}(0, 1)\},$$

wobei

$W = \mathbb{R}_{\geq 0}$  oder  $\mathbb{R}$  der Zustandsraum (Menge der möglichen Preise),  
 $X = \mathbb{R}$  die Menge der möglichen Innovationen,  
 $r : W \times X \rightarrow \mathbb{R}$  die unbehandelte Übergangsfunktion (r für „rough“),  
 $g : \mathbb{R} \rightarrow W = \mathbb{R}_{\geq 0}$  oder  $\mathbb{R}$  eine Glättungsfunktion,  
 $u : W \times X \rightarrow W$  die Übergangsfunktion des IFS.

Die Abbildung  $u$  geht dabei aus der Komposition von  $g$  und  $r$  hervor

$$W \times X \xrightarrow{r} \mathbb{R} \xrightarrow{g} W$$

$$u(w, x) = g(r(w, x)).$$

Gemäß der Überlegungen zu einem niedrigsten Wert  $\underline{w}_0 \in W = \mathbb{R}_{\geq 0}$  und einem höchsten Wert  $\overline{w}_0 \in W = \mathbb{R}_{\geq 0}$  ( $\underline{w}_0 < \overline{w}_0$ ), definiert man die Glättungsfunktion  $g : \mathbb{R} \rightarrow W$

$$g(w') = \begin{cases} \underline{w}_0 e^{-(\underline{w}_0 - w')} & , w' < \underline{w}_0 \\ w' & , \underline{w}_0 \leq w' \leq \overline{w}_0 \\ \overline{w}_0 + (1 - \gamma)(w' - \overline{w}_0) & , \overline{w}_0 < w' \end{cases}$$

$$0 < \gamma < 1.$$

Für den Fall  $W = \mathbb{R}$ ,  $-\infty < \underline{w}_0 < 0 < \overline{w}_0 < \infty$  verwendet man folgende Abwandlung

$$g(w') = \begin{cases} \underline{w}_0 - (1 - \alpha)(\underline{w}_0 - w') & , w' < \underline{w}_0 \\ w' & , \underline{w}_0 \leq w' \leq \overline{w}_0 \\ \overline{w}_0 + (1 - \gamma)(w' - \overline{w}_0) & , \overline{w}_0 < w' \end{cases}$$

$$0 < \alpha, \gamma < 1.$$

Den Einstieg in die Betrachtung bildet wiederum das AR(1)-ARCH(1)-Modell, welches als unbehandelte Übergangsfunktion dient.

Herkenrath [13] beschränkt die Betrachtung jedoch nicht nur auf den klassischen Fall, wie er in BuK [19] verwendet wird:

$$r(w, x) = \beta w + \sqrt{\alpha_0 + \alpha_1 w^2} x,$$

wobei  $\beta > 0$  konstant,  $\alpha_0 > 0$ ,  $\alpha_1 \geq 0$  und  $x$  als Realisation eines i.i.d.  $\mathcal{N}(0, 1)$ -Prozesses.

Vielmehr wird ein allgemeinerer Prozessaufbau durch die stetigen Funktionen  $\beta : W \rightarrow \mathbb{R}_{>0}$ ,  $k : W \rightarrow \mathbb{R}_{>0}$  konstruiert:

$$r(w, x) = \beta(w)w + k(w)x$$

und enthält den klassischen Ansatz als Spezialfall.

**Satz 2.1. (Theorem 2 - Herkenrath [13])**

*Unter den Bedingungen*

$$\beta(w) \leq 1 - \beta, \quad \text{mit } \beta > 0 \text{ für } w > \overline{w}_0,$$

$$k(w) \leq (1 - \zeta)w, \quad \text{mit } \zeta > 0 \text{ für } w > \overline{w}_0,$$

$$\lim_{w \rightarrow \infty} \frac{w}{k(w)} = l \quad \text{und} \quad (1 - \gamma)[(1 - \beta) + \frac{1 - \zeta}{\sqrt{2\pi}}] = 1 - \eta < 1,$$

*ist die durch das obige IFS (vgl.(I)) definierte Markov-Kette  $(W_t, t \in \mathbb{N}_0)$  geometrisch ergodisch.*

**Beweis zu Satz 2.1.**

Die Driftbedingung von Feigin und Tweedie [32], zusammen mit der Aufspaltung des Integrationsbereichs, ermöglicht dank der etwas kryptisch wirkenden Bedingungen eine geschickte Abschätzung und damit geometrische Ergodizität. Da die Details das Verständnis an dieser Stelle nicht voranbringen, sei dafür auf den Beweis zu Theorem 2 - [13] verwiesen.  $\square$

Weil in der Arbeit von Herkenrath [13] das AR(1)-ARCH(1)-Modell nicht direkt behandelt wird, sondern nur als Ausgangspunkt dient, sind die Ergebnisse insofern natürlich nicht uneingeschränkt miteinander vergleichbar. So bildet die eine Arbeit keine konsequente Fortführung und Erweiterung der anderen, aber es bietet sich damit zumindest die Möglichkeit, etwas Licht auf den ansonsten vielleicht undurchsichtigen „smoothing“-Effekt zu werfen. Vor dem Hintergrund der oben abgebildeten Tabellenwerte für  $\beta$  und  $\alpha_1$  aus BuK [19] und der Tatsache, dass dort genauso wie in Herkenrath [13] die  $\mathcal{N}(0, 1)$ -Verteilung dem Störterm zugrunde liegt, fallen einem sofort einige wichtige Unterschiede ins Auge.

In BuK [19] wird der zulässige Bereich für  $\beta$  dahingehend erweitert, dass nicht unbedingt  $0 < \beta < 1$  gelten muss um geometrische Ergodizität folgern zu können. Mit welchem Aufwand die Erweiterung auf  $\beta > 1$  erkaufte wird, zeigen die tabellierten Werte für  $\alpha_1$  bzw.  $\beta$  oder noch eindrucksvoller die Veranschaulichung ihrer Entstehung im Anhang (Abschnitt 4.3). Bei Herkenrath [13] hingegen entsteht durch die erste Bedingung (Satz 2.1) die Forderung  $0 < \beta(w) < 1$ , jedoch entscheidender Weise nur für  $w > \overline{w}_0$ , d.h. womöglich weit außerhalb des interessierenden Bereichs. Die besonders großen Werte sind nicht nur für praktische Betrachtungen meist uninteressant, sondern auch für die Beweisführung kommt ihnen oft eine gesonderte Rolle zu - vgl. Feigin und Tweedie [32] und die Glättungsidee.

Für einen weiteren Vergleich der Ergebnisse, ist es zunächst notwendig, eine genauere Vorstellung vom Zustandekommen der in den obigen Tabellen festgehaltenen Zahlenwerte zu gewinnen.

In einem ersten Schritt, kann man die Bedingung

$$h'_{\beta, \alpha_1}(0) = \mathbb{E}(\log |\beta + \sqrt{\alpha_1} \varepsilon|) < 0$$

mit Hilfe der Jensenschen Ungleichung ( $\log \sim$  konkav) nach oben abschätzen

$$\begin{aligned} \mathbb{E}(\log |\beta + \sqrt{\alpha_1} \varepsilon|) &\leq \log \mathbb{E}(|\beta + \sqrt{\alpha_1} \varepsilon|) \\ &\leq \log \mathbb{E}(|\beta| + |\sqrt{\alpha_1} \varepsilon|) \\ &= \log \mathbb{E}(|\beta| + |\sqrt{\alpha_1}| |\varepsilon|) \\ &= \log(|\beta| + |\sqrt{\alpha_1}| \mathbb{E}(|\varepsilon|)) \end{aligned}$$

und erkennt so, dass bei einer geeigneten Wahl der Parameter  $\beta, \alpha_1$  die Bedingung

$$|\beta| + |\sqrt{\alpha_1}| \mathbb{E}(|\varepsilon|) < 1$$

das Geforderte sicherstellt. Dies bedeutet insbesondere aber, dass  $|\beta| < 1$  gewählt werden muss. Die obigen Tabellen geben jedoch einen viel größeren Zulässigkeitsbereich für  $\beta$  und  $\alpha_1$  an, als es auf Grundlage der Jensenschen Ungleichung möglich wäre.

Die konkreten Zahlenwerte ergeben sich unter Zuhilfenahme eines Iterationsverfahrens, und zwar wird das Newton-Iterations-Verfahren verwendet, um Lösungen der Gleichung

$$\mathbb{E}(\log |\beta + \sqrt{\alpha_1} \varepsilon|) \stackrel{!}{=} 0$$

zu erhalten. Entweder wird ein  $\beta$  festgehalten und ein  $\alpha_1$  mit jedem Iterationsvorgang verbessert oder genau umgekehrt. Den Programmcode für Maple findet man in Kiefersbeck [17], genauso wie weitere Erläuterungen zur Rechengenauigkeit und dem Konvergenzverhalten, da sich nur bei bereits guten Startwerten eine Lösung findet. Für die Laplace-Verteilung und die Student-t-Verteilung werden ganz analoge Programme angegeben - vgl. Anhang von [17].

Die Ergebnisse von BuK [19] sind zwar nur computergestützt berechenbar, jedoch zumindest durch eine Plausibilitätsüberlegung (vgl. Anhang 4.3) nachvollziehbar und bieten durch einen Vergleich zum geglätteten Ansatz die Möglichkeit, die dadurch hervorgerufenen Effekte zu bewerten.

## 2.1 Smoothing bewahrt die Kontraktionsbedingung

Aus den oben tabellierten Werten einerseits und dem Grundgedanken des „smoothing“ sowie den hervorgebrachten Ergebnissen andererseits, entstehen ganz natürlich einige interessante Querverbindungen und Fragen, die es zu beantworten gilt. Insbesondere der Fall  $\beta > 1$ , der ja bei BuK auch zu geometrischer Ergodizität genügt, gibt Anlass zu der Vermutung, dass es sich im „geglätteten“-Beispiel vielleicht ähnlich verhält. Ganz salopp formuliert, ergibt sich die folgende Frage:

Wenn  $r(w, x)$  die Kontraktionsbedingung erfüllt, so geht diese Eigenschaft durch „smoothing“ verloren oder genügt der „geglättete“ Prozess dann auch einer Kontraktionsbedingung!?

Wählt man eine Parameterkonstellation aus BuK (gemäß obiger Tabellen), so hat der Ausgangsprozess, dann wie gesehen, die Eigenschaft „Kontraktiv im Mittel“ und der Effekt einer überlagerten Glättungsfunktion (wie etwa)

$$g(w') = \begin{cases} \underline{w}_0 e^{-(\underline{w}_0 - w')} & , w' < \underline{w}_0 \\ w' & , \underline{w}_0 \leq w' \leq \overline{w}_0 \\ \overline{w}_0 + (w' - \overline{w}_0)^\alpha & , \overline{w}_0 < w' \end{cases} \quad 0 < \alpha < 1$$

ergibt sich nun wie folgt:

Innerhalb eines Kompaktums  $[\underline{w}_0, \overline{w}_0]$  bleibt der Prozess unverändert. Dieser Bereich ergibt sich, wie eingangs besprochen, aus praktischen Überlegungen bzw. Betrachtungen des Alltagsgeschehens. Außerhalb dieses Bereichs wird der Prozess gemäß obiger Funktionsvorschrift „geglättet“ oder ganz anschaulich „zusammengestaucht“. Der Grundgedanke des Glättungsansatzes ist nun der, dass man innerhalb eines Kompaktums wenig von dem Prozess fordert, etwa nur Stetigkeit und außerhalb künstlich eine Kontraktion hervorruft. Wie zu Beginn dieses Abschnitts gesehen, ist man stets bemüht, die geforderten Bedingungen zu verallgemeinern oder abzuschwächen und eben diese Zielsetzung verfolgt man mit dem Glättungsansatz auch. Man möchte also die ansonsten globale Kontraktionsbedingung abschwächen, indem man nur außerhalb eines Kompaktums die Kontraktionseigenschaft fordert.

Wenn man die obige Glättungsfunktion betrachtet und sich Gedanken darüber macht, ob denn der neue „geglättete“ Prozess tatsächlich eine Kontraktionsbedingung erfüllt, weil es der Ausgangsprozess tut, begegnen einem doch gleich einige Stolpersteine. So bildet die Glättungsfunktion  $g(w')$  deutlich außerhalb des Kompaktums  $[\underline{w}_0, \overline{w}_0]$  eine Kontraktion, aber unmittelbar an den Nahtstellen zu diesem eben nicht. Dieser Umstand behindert einen doch sehr in der Argumentation, wird im Anhang (vgl. Abschnitt 4.4) veranschaulicht und führt zu der stärkeren Glättungsfunktion:

$$g^*(w') = \begin{cases} e^{-(\underline{w}_0 - w')} + (\underline{w}_0 - 1) & , w' < \underline{w}_0, \quad \underline{w}_0 \geq 1 \\ w' & , \underline{w}_0 \leq w' \leq \overline{w}_0 \\ (\overline{w}_0 - 1) + (w' - \overline{w}_0 + 1)^\alpha & , \overline{w}_0 < w' \end{cases} \quad 0 < \alpha < 1$$

Die Konsequenz dieser veränderten Glättungsfunktion verdeutlicht der folgende Satz:

**Satz 2.2.**

Sei  $u : (W \times X, \mathcal{B}_W \otimes \mathcal{X}) \rightarrow (W, \mathcal{B}_W)$  die Übergangsfunktion eines IFS (einer SRE) und  $u_t = u(\cdot, X_t)$  erfülle die Bedingung „contractive on average“ („Kontraktiv im Mittel“)

$$|u_t(w) - u_t(\hat{w})| \leq \Lambda(u_t)|w - \hat{w}| \quad \text{und} \quad \mathbb{E}[\log \Lambda(u_0)] < 0.$$

Durch Komposition von  $u$  mit einer Glättungsfunktion

$$g^*(w') = \begin{cases} e^{-(\underline{w}_0 - w')} + (\underline{w}_0 - 1) & , w' < \underline{w}_0, \quad \underline{w}_0 \geq 1 \\ w' & , \underline{w}_0 \leq w' \leq \overline{w}_0 \\ (\overline{w}_0 - 1) + (w' - \overline{w}_0 + 1)^\alpha & , \overline{w}_0 < w' \quad 0 < \alpha < 1 \end{cases}$$

entsteht eine neue Übergangsfunktion  $u_t^* = g^*(u_t(\cdot, X_t))$  und die Eigenschaft „contractive on average“ bleibt erhalten.

Wenn man nun den Gesamtprozess auf seine Eigenschaften hin untersuchen will, so erscheint es zunächst einmal plausibel, die Eigenschaften des Ausgangsprozesses als Grundlage heranzuziehen und aufbauend darauf seine Untersuchung durchzuführen. Um etwas über das Kontraktionsverhalten des Gesamtprozesses zu erfahren, wird der Lipschitzkoeffizient betrachtet, um dann aus dem Zusammenhang zum Ausgangsprozess Schlussfolgerungen zu ziehen. Für den Beweis nimmt man exemplarisch einen AR(1)-ARCH(1)-Ausgangsprozess an, der die Eigenschaft „Kontraktiv im Mittel“ erfüllt.

**Beweis zu Satz 2.2.**

Der Lipschitzkoeffizient des geglätteten Prozesses, ergibt sich zu:

$$\begin{aligned} \Lambda(u_t^*) &= \sup_{\substack{w, \hat{w} \in W \\ w \neq \hat{w}}} \frac{|g_t^*(r(w, x)) - g_t^*(r(\hat{w}, x))|}{|w - \hat{w}|} \\ &= \sup_{\substack{w, \hat{w} \in W \\ w \neq \hat{w}}} \frac{|g_t^*(\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} x) - g_t^*(\beta \hat{w} + \sqrt{\alpha_0 + \alpha_1 \hat{w}^2} x)|}{|w - \hat{w}|} \end{aligned}$$

Mittels einer Fallunterscheidung hat man nun die Möglichkeit die bisherigen Ergebnisse direkt auszunutzen. Es interessieren dabei die unterschiedlichen Konstellationen in denen

$$\dot{w} := \beta w + \sqrt{\alpha_0 + \alpha_1 w^2} x = r(w, x) \quad \text{und} \quad \ddot{w} := \beta \hat{w} + \sqrt{\alpha_0 + \alpha_1 \hat{w}^2} x = r(\hat{w}, x)$$

auftauchen können (in Bezug zu dem vorgegebenen Kompaktum).

- 1. Fall** Liegen  $\dot{w}$  und  $\ddot{w}$  im Kompaktum  $[\underline{w}_0, \overline{w}_0]$ , so bedeutet das in Hinblick auf den Wert des Supremums keine Veränderung. Das Verhalten des Prozesses stimmt in diesem Bereich ja mit dem des Ausgangsprozesses überein.
- 2. Fall** Gilt  $\dot{w}, \ddot{w} < \underline{w}_0$ , so folgt wegen der Kontraktionseigenschaft der Glättungsfunktion insbesondere  $|g^*(\dot{w}) - g^*(\ddot{w})| < |\dot{w} - \ddot{w}|$  und damit kleinere Werte für die Supremumsbildung, als es im Ausgangsprozess der Fall war.

- 3. Fall** Gilt  $\dot{w}, \ddot{w} > \overline{w_0}$ , so folgt wegen der Kontraktionseigenschaft der Glättungsfunktion insbesondere  $|g^*(\dot{w}) - g^*(\ddot{w})| < |\dot{w} - \ddot{w}|$  und damit kleinere Werte für die Supremumsbildung, als es im Ausgangsprozess der Fall war.
- 4. Fall** O.B.d.A. sei  $\dot{w} < \ddot{w}$ . Für den Fall  $\dot{w} < \underline{w_0}$  und  $\ddot{w} \in [\underline{w_0}, \overline{w_0}]$  ergibt sich wegen der Stauchungseigenschaft der Glättungsfunktion (Bildmenge oberhalb der Identität - siehe Anhang 4.4) ein Wert  $\dot{w} < g^*(\dot{w}) = \ddot{w} < \underline{w_0}$ . Damit folgt dann  $|g^*(\dot{w}) - g^*(\ddot{w})| < |\dot{w} - \ddot{w}|$  und wiederum kleinere Werte für die Supremumsbildung als beim Ausgangsprozess.
- 5. Fall** O.B.d.A. sei  $\dot{w} < \ddot{w}$ . Für den Fall  $\dot{w} \in [\underline{w_0}, \overline{w_0}]$  und  $\ddot{w} > \overline{w_0}$  ergibt sich wegen der Stauchungseigenschaft der Glättungsfunktion (Bildmenge unterhalb der Identität - siehe Anhang 4.4) ein Wert  $\overline{w_0} < g^*(\dot{w}) = \ddot{w} < \ddot{w}$ . Damit folgt dann  $|g^*(\dot{w}) - g^*(\ddot{w})| < |\dot{w} - \ddot{w}|$  und wiederum kleinere Werte für die Supremumsbildung als beim Ausgangsprozess.
- 6. Fall** O.B.d.A. sei  $\dot{w} < \ddot{w}$ . Die einzige noch nicht abgehandelte Konstellation ist  $\dot{w} < \underline{w_0}$  und  $\overline{w_0} < \ddot{w}$ . Eine Kombination aus Fall 4 und 5 kommt hier zum tragen, nur mit dem Unterschied, dass von beiden Seiten aus gestaucht wird. Insgesamt folgt wie oben auch  $|g^*(\dot{w}) - g^*(\ddot{w})| < |\dot{w} - \ddot{w}|$ .

Es bleibt festzuhalten

$$\Lambda(u_0^*) = \sup_{\substack{w, \hat{w} \in W \\ w \neq \hat{w}}} \frac{|g^*(r(w, x)) - g^*(r(\hat{w}, x))|}{|w - \hat{w}|} \leq \sup_{\substack{w, \hat{w} \in W \\ w \neq \hat{w}}} \frac{|r(w, x) - r(\hat{w}, x)|}{|w - \hat{w}|} = \Lambda(r_0)$$

und da die Parameter ja so gewählt waren, dass  $\mathbb{E}[\log \Lambda(r_0)] < 0$  „contractive on average“ erfüllt wird, ergibt sich demnach die gleiche Eigenschaft auch für den neuen - geglätteten Prozess  $\mathbb{E}[\log \Lambda(u_0^*)] < 0$ .  $\square$

Mit der stärkeren Glättungsfunktion  $g^*(\cdot)$  kann man die Eigenschaft „contractive on average“ also vom Ausgangsprozess zum geglätteten Gesamtprozess hinüberretten. Bemerkenswert ist an dieser Stelle auch, dass dies völlig losgelöst von sonstigen Modellcharakteristika geschieht (AR(1)-ARCH(1) dient nur der Veranschaulichung) und auch das stärkere Konzept der Lipschitz-Kontraktivität ( $\Lambda(r_0) < 1$ ) lässt sich analog überführen.

### 2.1.1 Glättung mittels Geradenstücken

Die für die obigen Schlussfolgerungen benötigten Kontraktions- und Stauchungseigenschaften ließen sich auch mit einfacheren Funktionen erreichen: Eine Glättungsfunktion in Form einer Geraden mit einer Steigung zwischen 0 und 1 liefert die Kontraktionseigenschaft außerhalb des Kompaktums. Für die Stauchungseigenschaft und aus Stetigkeitsgründen muss eine solche Gerade dann nur durch  $\underline{w_0}$  bzw.  $\overline{w_0}$  gehen, um die Eigenschaft „contractive on average“ zu bewahren.

Eine etwas ausführlichere Behandlung und Veranschaulichung der genannten Eigenschaften bzw. der dadurch zulässigen Glättungsfunktionen, geschieht im Anschluss an eine Rekapitulation und Sammlung der bisherigen Ergebnisse, in Form von Sätzen und Lemmata.

**Satz 2.3.**

Sei  $u : (W \times X, \mathcal{B}_W \otimes \mathcal{X}) \rightarrow (W, \mathcal{B}_W)$  die Übergangsfunktion eines IFS (einer SRE) und  $u_t = u(\cdot, X_t)$  erfülle die Bedingung „contractive on average“ („Kontraktiv im Mittel“)

$$|u_t(w) - u_t(\hat{w})| \leq \Lambda(u_t)|w - \hat{w}| \quad \text{und} \quad \mathbb{E}[\log \Lambda(u_0)] < 0.$$

Durch Komposition von  $u$  mit einer Glättungsfunktion

$$g^{**}(w') = \begin{cases} \frac{w_0}{w'} + \alpha(w' - \frac{w_0}{w'}) & , w' < \frac{w_0}{w_0} \\ \frac{w_0}{w'} & , \frac{w_0}{w_0} \leq w' \leq \frac{w_0}{w_0} \\ \frac{w_0}{w'} + \gamma(w' - \frac{w_0}{w_0}) & , \frac{w_0}{w_0} < w' \end{cases} \quad 0 < \alpha, \gamma < 1$$

entsteht eine neue Übergangsfunktion  $u_t^{**} = g^{**}(u_t(\cdot, X_t))$  und die Eigenschaft „contractive on average“ bleibt erhalten.

Hierbei können  $\alpha$  und  $\gamma$  unabhängig voneinander gewählt werden ( $0 < \alpha, \gamma < 1$ ).

**Beweis zu Satz 2.3.**

Analog zum Beweis von Satz 2.2. Die Stauchungs- und Kontraktionseigenschaft außerhalb des Kompaktums, ergeben sich in diesem Fall noch offensichtlicher.  $\square$

Mit Satz 2.2., 2.3. ist zumindest die bereits mehrfach erwähnte Hauptbedingung für die Existenz einer stationären Anfangsverteilung gesichert. Es stellt sich nun die Frage, ob denn auch die restlichen Eigenschaften erhalten bleiben und man analoge Aussagen über den neuen Gesamtprozess beweisen kann. Um jetzt von dem konkreten Beispiel etwas wegzukommen und allgemein gültige Aussagen zu gewinnen, betrachte man das oben bereits erwähnte Theorem 3.3 von Iosifescu.

**Satz 2.4. (Theorem 3.3 - S. 11 Iosifescu [16])**

Sei  $(W, d)$  ein vollständig separabler metrischer Raum. Wenn weiter (2.4) und (2.5) erfüllt sind, so besitzt die Markov-Kette  $(W_n, n \in \mathbb{N})$  eine einzige stationäre Anfangsverteilung  $\pi$  und

$$\rho_L(P^n(w, \cdot), \pi) \leq \frac{l^n}{1-l} \int_X d(w, u_x(w)) p(dx)$$

für  $n \in \mathbb{N}$  und  $w \in W$  beliebig. Über  $(\Omega, \mathcal{K}, \mathbb{P}_{\pi,p})$  ist die Folge  $(W_n, n \in \mathbb{N})$  ein ergodischer, streng stationärer Prozess.

Die Bedingung (2.4) ist bereits abgehakt (s.o.) und für eine analoge Aussage bezüglich des geglätteten Prozesses, muss demnach nur die Gültigkeit von (2.5) gezeigt werden:

**Satz 2.5.**

Der durch die aufgesetzte Glättungsfunktion entstandene Gesamtprozess, erfüllt die Bedingung

$$\int_X d(w_0, g^*(u_x(w_0))) p(dx) < \infty$$

wenn der Ausgangsprozess die Bedingung

$$\int_X d(w_0, u_x(w_0)) p(dx) < \infty \quad (2.5)$$

erfüllt.

**Beweis zu Satz 2.5.**

Es gilt  $g^* \in \text{Lip}_1(W) = \{f : W \rightarrow \mathbb{R} \mid s(f) \leq 1\}^2$ , wobei

$$s(f) = s(f, d) := \sup_{\substack{w' \neq w'' \\ w', w'' \in W}} \frac{|f(w') - f(w'')|}{d(w', w'')}$$

und damit ergibt sich

$$|g^*(u_x(w))| \leq |g^*(w)| + d(w, u_x(w)), \quad x \in X, \quad w \in W.$$

Aus der Bedingung (2.5) und der Definition<sup>3</sup> des Übergangsoperators  $U$ , folgt dann

$$|Ug^*(w)| \leq \int_X |g^*(u_x(w))| p(dx) < \infty, \quad w \in W. \quad (*^1)$$

Mit der Dreiecksungleichung ergibt sich zunächst

$$\begin{aligned} \int_X d(w_0, g^*(u_x(w_0))) p(dx) &\leq \int_X d(w_0, u_x(w_0)) p(dx) + \int_X d(u_x(w_0), g^*(u_x(w_0))) p(dx) \\ &\leq \int_X d(w_0, u_x(w_0)) p(dx) + \int_X |u_x(w_0)| p(dx) + \int_X |g^*(u_x(w_0))| p(dx) \end{aligned}$$

mit dem vorangegangenen

$$\begin{aligned} &= \stackrel{(*^2)}{< \infty} + \stackrel{(*^3)}{< \infty} + \stackrel{(*^4)}{< \infty} \\ &< \infty \end{aligned}$$

und somit die gewünschte Aussage.

Hierbei folgt  $(*^2)$  aus der Bedingung (2.5),  $(*^3)$  ergibt sich durch  $(*^1)$ , die Festlegung  $f = g^* = \text{Id}$  und  $(*^4)$  ist eine direkte Anwendung von  $(*^1)$ .  $\square$

---

<sup>2</sup>vgl. Iosifescu - APPENDIX 1

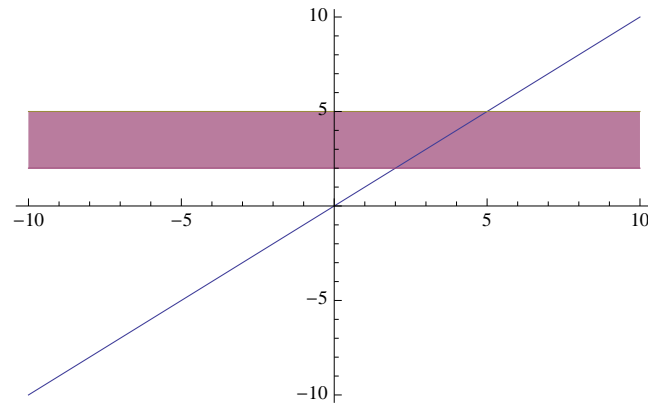
<sup>3</sup>vgl. Iosifescu - (2.5), (2.3) und S.8 (oben)



Damit sind alle Bedingungen von Satz 2.4. erfüllt und die Existenz einer stationären Anfangsverteilung für den geglätteten Gesamtprozess gesichert.

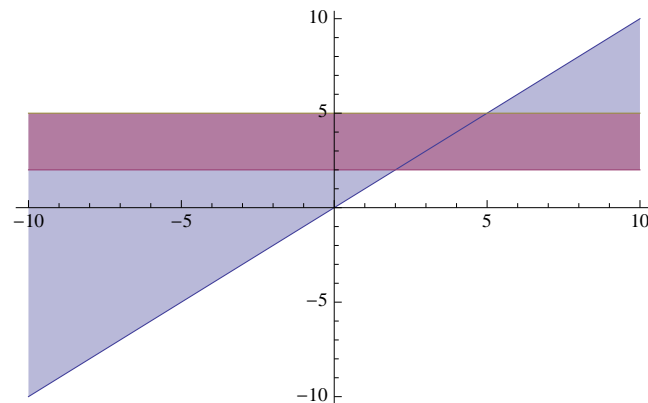
Um nun eine genauere Vorstellung von den zulässigen Glättungsfunktionen in den obigen Anwendungen zu gewinnen, wird das konkrete Kompaktum  $[\underline{w}_0, \overline{w}_0] = [2, 5]$  als Anschauungsbeispiel herangezogen (vgl. Abschnitt 4.4).

Die Identität als triviale Glättungsfunktion bildet den Startpunkt der weiteren Betrachtung - vgl. Konstruktion von  $g^*$  bzw.  $g^{**}$  oben.

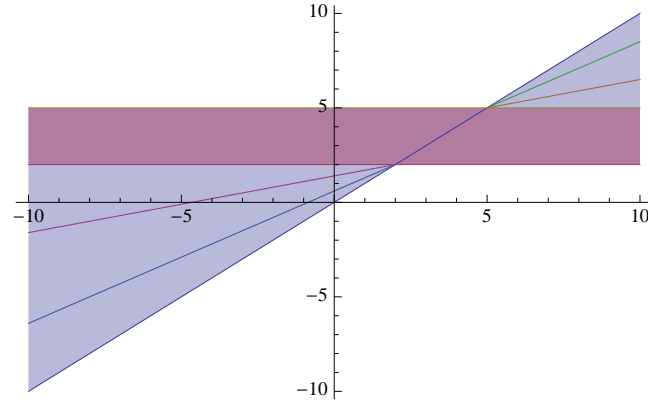


Der rötlich markierte Bereich, steht für das durch vorangegangene Überlegungen zu definierende Kompaktum und zur Interpretation der Grafik gibt es folgendes zu sagen: Die x-Achse bildet den möglichen Wertebereich des Ausgangsprozesses und die durch die (blaue) Glättungsfunktion hervorgebrachten y-Werte bilden den Glättungseffekt ab (in obiger Grafik nur Identität).

Um einen Glättunseffekt außerhalb des Kompaktums zu erreichen, ist die Identität eine triviale Beschränkung (s.o.) und man erhält als zulässigen Bereich für die Glättungsfunktion die blau markierten Flächen:



Es sollen an dieser Stelle also erst einmal nur Glättungsfunktionen betrachtet werden, die über dem Kompaktum der Identität entsprechen und außerhalb davon die blauen Flächen nicht verlassen. Die einfachsten Beispiele für diese Konstruktion wurden oben bereits erwähnt und bieten, neben der einfachen Konstruktion, auch noch weitere Vorteile.



Die aus Geradenstücken zusammengesetzten Glättungsfunktionen, wie oben angedeutet, haben den großen Vorteil, dass sie sehr deutlich eine Art Stellschraube für den Grad der Glättung bereitstellen. Je nach Winkel oder Steigung der Geraden, die an die Randpunkte des Kompaktums  $[w_0, \overline{w_0}]$  anschließen, kann man den Glättungseffekt verstärken oder abschwächen und somit eine problemangepasste Glättung vornehmen.

Für die oben geforderten Kontraktions- und Stauchungseigenschaft reicht es aber nicht aus, dass sich die Glättungsfunktion im blauen Bereich bewegt. Vielmehr kommt es darauf an, dass Sie zusätzlich eine Steigung  $\leq 1$  besitzt, um sich Ausnahmefälle wie zu Beginn der Konstruktion von  $g$  und  $g^*$  zu ersparen.

## 2.2 Driftbedingung - geometrische Ergodizität

Mit Satz 2.2. bis Satz 2.5. wurde gezeigt, dass sich die Kontraktionseigenschaft und damit die Existenz einer stationären Anfangsverteilung vom ungeglätteten Prozess zum geglätteten Prozess übertragen lässt. Da für die geometrische Ergodizität des AR(1)-ARCH(1)-Prozesses, wie oben gesehen, sowohl im geglätteten, als auch im ungeglätteten Fall die Driftbedingung von Feigin und Tweedie [32] genutzt wurde, steht im Rahmen der bisherigen Vorgehensweise folgende Frage im Raum:

**Wenn der Ausgangsprozess  $\{W_t\}$  die Driftbedingung erfüllt, so geht diese Eigenschaft durch „smoothing“ verloren oder genügt der „geglättete“ Prozess  $\{\tilde{W}_t\}$  dann auch einer Driftbedingung?**

Um sich dieser Frage zu nähern, bedarf es erst einmal einer genaueren Vorstellung darüber, was die Driftbedingung bedeutet und unter welchen Voraussetzungen sie in beiden Fällen (ungeglättet und geglättet) gezeigt wurde.

**Satz 2.6. (Theorem 1 - [32])**

Seien  $\{W_t, t \in \mathbb{N}_0\}$  eine Feller-Kette,  $\phi$  ein Maß und  $A$  eine kompakte Menge mit  $\phi(A) > 0$ , so dass

(i)  $\{W_t, t \in \mathbb{N}_0\}$  ist  $\phi$ -irreduzibel,

(ii) es existiert eine nicht-negative, stetige Funktion  $h : W \rightarrow \mathbb{R}$  mit

$$h(w) \geq 1 \quad \text{für } w \in A \quad (2.1)$$

und für ein  $\delta > 0$

$$\mathbb{E}[h(W_t) \mid W_{t-1} = w] \leq (1 - \delta)h(w), \quad \text{für } w \in A^c \text{ (Komplement von } A). \quad (2.2)$$

Dann ist  $\{W_t, t \in \mathbb{N}_0\}$  geometrisch ergodisch.

Bedingungen, um diesem Theorem zu genügen, stellen sich in BuK [19] für ein AR(1)-Modell mit ARCH(1)-Fehler/Innovation wie folgt dar:

$$W_t = \beta W_{t-1} + \sqrt{\alpha_0 + \alpha_1 W_{t-1}^2} \varepsilon_t, \quad t \in \mathbb{N},$$

wobei  $\beta \in \mathbb{R}$ ,  $\alpha_0 > 0$ ,  $\alpha_1 \geq 0$ ,  $\{\varepsilon_t, t \in \mathbb{N}\}$  i.i.d. mit Erwartungswert 0 und Varianz 1. Weiter sei  $W_0$  unabhängig von  $\{\varepsilon_t, t \in \mathbb{N}\}$ .

(B1) Der Störterm  $\{\varepsilon_t, t \in \mathbb{N}\}$  besitze eine symmetrische, strikt positive und stetige Lebesgue-Dichte über  $(-\infty, +\infty)$ , sowie ein existentes zweites Moment  $\mathbb{E}(\varepsilon_t^2) < \infty$ .

(B2) Der Parameterraum sei

$$\Theta = \{\theta = (\beta, \alpha_0, \alpha_1)^T : \mathbb{E}(\log |\beta + \sqrt{\alpha_1} \varepsilon_t|) < 0, -\infty < \beta < +\infty, \alpha_0 > 0, \alpha_1 \geq 0\}$$

Die Bedeutung und Gestalt der Hauptbedingung  $\mathbb{E}(\log |\beta + \sqrt{\alpha_1} \varepsilon_t|) < 0$ , wurde oben ja bereits ausführlich thematisiert. Ebenso die Tatsache, dass die Wahl der Standardnormalverteilung als Störterm  $\varepsilon \sim \mathcal{N}(0, 1)$  die oben besprochenen Werte hervorbringt. Der unterschiedliche zulässige Parameterbereich im ungeglätteten und geglätteten Fall des AR(1)-ARCH(1)-Modells, war ja ausschlaggebend für die weitere Betrachtung und so soll das Bestreben jetzt darin liegen, diese Unterschiede nach Möglichkeit zu überwinden.

Die in Satz 2.6. auftauchende Funktion  $h$  wird auch als „Drift“- oder „Testfunktion“ bezeichnet und ihre genaue Festlegung oder geschickte Wahl im Zusammenhang mit geometrischer Ergodizität als eine Art „Kunst“ angesehen (siehe z.B. [34]).

Für die weitere Betrachtung wird die Wahl von  $h$  in Borkovec und Klüppelberg [19]

$$h(w) = 1 + |w|^\eta, \quad \eta \in (0, \min(K, 2))$$

zugrunde gelegt. Die genaue Höhe von  $K$  entspringt dabei dem Beweis in BuK [19], spielt für unsere Überlegungen im weiteren aber keine Rolle, da auch der Umstand reichen wird, dass  $\eta$  in  $(0, 2)$  ist. Etwas wichtiger für die Driftbedingung, ist die Wahl von  $A = [-M, M]$  und damit von  $A^c = \{w \mid |w| > M\}$ , für  $M$  hinreichend groß.

**Satz 2.7.**

Wenn der  $AR(1)$ - $ARCH(1)$ -Prozess  $\{W_t\}$ , unter den Voraussetzungen von BuK [19] die Driftbedingung erfüllt, so tut es auch der durch Glättung entstandene Prozess  $\{\tilde{W}_t\}$ .

Mit Satz 2.6. erhält man für die Konstellation in BuK [19]

$$\mathbb{E}[h(W_t) \mid W_{t-1} = w] \leq (1 - \delta)h(w), \quad \text{für } w \in A^c$$

oder genauer

$$\begin{aligned} & \mathbb{E}[h(\beta W_{t-1} + \sqrt{\alpha_0 + \alpha_1 W_{t-1}^2} \varepsilon_t) \mid W_{t-1} = w] \\ &= \mathbb{E}[h(\underbrace{\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t}_{(*)})] \leq (1 - \delta)h(w). \end{aligned}$$

Betrachtet man jetzt den geglätteten Prozess  $\tilde{W}_t$ , der immer nach folgendem Schema entsteht

$$\begin{array}{ccccccc} W_{t-1} & & W_t & & W_{t+1} & & \cdots \\ \downarrow g^{**} & \nearrow & \downarrow g^{**} & \nearrow & \downarrow g^{**} & \nearrow & \\ \tilde{W}_{t-1} & & \tilde{W}_t & & \tilde{W}_{t+1}, & & \end{array}$$

so erhält man in völliger Analogie

$$\begin{aligned} & \mathbb{E}[h(\tilde{W}_t) \mid \tilde{W}_{t-1} = w] \\ &= \mathbb{E}[h(g^{**}(\beta \tilde{W}_{t-1} + \sqrt{\alpha_0 + \alpha_1 \tilde{W}_{t-1}^2} \varepsilon_t)) \mid \tilde{W}_{t-1} = w] \\ &= \mathbb{E}[h(g^{**}(\underbrace{\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t}_{(*)}))]. \end{aligned}$$

Zunächst lässt sich festhalten, dass in beiden Fällen  $(*)$  auftaucht und bei der trivialen Wahl  $g^{**} = \text{Id}$  als Glättungsfunktion, sogar kein Unterschied vorhanden wäre.

Des Weiteren gilt  $(*) \sim \mathcal{N}(\beta w, \alpha_0 + \alpha_1 w^2)$  und mit dem Transformationssatz für Dichten wird nun gezeigt, dass sich durch die aufgesetzte Glättungsfunktion die Situation zwar verändert, aber immer noch eine Abschätzung möglich ist, wie in Theorem 1 gefordert. Mit anderen Worten - der geglättete Prozess ist geometrisch ergodisch, wenn es der ungeglättete Prozess bereits war.

**Satz 2.8. Transformationssatz für Dichten** (Satz 8.5 [14] bzw. Satz 1.101 [18])

Sei  $X : \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable mit stetiger Verteilungsfunktion  $F_X$  und Dichte  $f_X(x) = \frac{\partial F(x)}{\partial x}$  bezüglich des Lebesgue-Maßes  $\lambda$ . Weiterhin sei  $g : \mathbb{R} \rightarrow \mathbb{R}$  bijektiv und stetig differenzierbar,  $\frac{\partial g(x)}{\partial x} \neq 0$ .

$\Rightarrow g \circ X$  hat die Dichte  $f_{g \circ X}(y) = f_X(g^{-1}(y)) \cdot \left| \frac{\partial g^{-1}(y)}{\partial y} \right|$  bezüglich  $\lambda$ .

Für eine stückweise stetige Funktion gilt:

Sei  $\mathbb{R} \supset I = \bigcup_{m \in M} I_m$  eine disjunkte Vereinigung offener Intervalle ( $I_m = ]a_m, b_m[$ ) derart, dass gilt:

a)  $F_X$  ist stetig differenzierbar auf  $I$ :

$$f_X(x) = \frac{\partial F_X(x)}{\partial x}$$

b)  $\mathbb{P}(X \in I) = 1$

c) Sei  $g_m = g|_{I_m}$ ,  $g_m : I_m \rightarrow \mathbb{R}$  die Einschränkung von  $g$  auf  $I_m$  mit

(i)  $g_m$  ist stetig differenzierbar und bijektiv

(ii)  $\frac{\partial g_m(x)}{\partial x} \neq 0$

Dann gilt:  $g \circ X$  hat die Dichte

$$f_{g \circ X}(y) = \sum_{m \in M} v_m(y),$$

$$\text{wobei } v_m(y) = \begin{cases} f_X(g_m^{-1}(y)) \cdot \left| \frac{\partial g_m^{-1}(y)}{\partial y} \right| & y \in I_m \\ 0 & \text{sonst.} \end{cases}$$

Wie bereits gesehen, hat  $\varepsilon_t \sim \mathcal{N}(0, 1)$  zu Folge  $(*) \sim \mathcal{N}(\beta w, \alpha_0 + \alpha_1 w^2)$ . Ab jetzt sei  $X$  eine Zufallsvariable mit  $X \sim \mathcal{N}(\beta w, \alpha_0 + \alpha_1 w^2)$  und der Dichte

$$f_X(x) = \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(x - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}}$$

Mit der Festlegung von  $h(x)$  und  $(*)$ , stellt sich das Ergebnis von Borkovec und Klüppelberg nun wie folgt dar:

$$1 + \int_{-\infty}^{+\infty} |x|^\eta f_X(x) dx \leq (1 - \delta) h(w) \quad (5)$$

Das in (5) auftauchende Integral wollen wir etwas umformen und so in Verbindung setzen zum geglätteten Fall. Entweder durch Anwendung des Transformationssatzes oder durch Überlegungen zur Auswirkung der Betragsfunktion auf die Dichte (vgl. Anhang 4.3), ergibt sich

$$\int_{-\infty}^{+\infty} |x|^\eta f_X(x) dx = \int_0^{+\infty} x^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} \left( e^{-\frac{(x - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} + e^{-\frac{((-x) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \right) dx. \quad (6)$$

Die Darstellung in (6) kann auch als Summe zweier Integrale aufgefasst werden

$$(6) = \int_0^{+\infty} x^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(x-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx + \int_0^{+\infty} x^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((-x)-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx$$

und jedes einzelne dieser beiden Integrale dann wiederum in drei Teilintegrale aufgespalten werden. Die Aufspaltung wird in der Art vollzogen, dass man den Integrationsbereich gemäß dem Definitionsbereich der Glättungsfunktion aufteilt: (Unterhalb des Kompaktums, das Kompaktum  $[\underline{w}_0, \overline{w}_0]$  selbst, oberhalb des Kompaktums).

Insgesamt erhält man für das Integral in (5) bzw. (6) die Darstellung:

$$(6) = \int_0^{\underline{w}_0} x^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(x-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx \quad (i)$$

$$+ \int_{\underline{w}_0}^{\overline{w}_0} x^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(x-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx \quad (ii)$$

$$+ \int_{\overline{w}_0}^{+\infty} x^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(x-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx \quad (iii)$$

$$+ \int_0^{\underline{w}_0} x^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((-x)-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx \quad (iv)$$

$$+ \int_{\underline{w}_0}^{\overline{w}_0} x^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((-x)-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx \quad (v)$$

$$+ \int_{\overline{w}_0}^{+\infty} x^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((-x)-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx \quad (vi)$$

Diese Darstellung mag im Moment noch etwas ungewöhnlich aussehen, ist allerdings angepasst an die durch die Glättungsfunktion und die Anwendung des Transformationssatzes entstehende Gestalt und bietet somit die beste Vergleichsmöglichkeit.

**Beweis zu Satz 2.7.**

In Teilen des Beweises wird auf den Anhang verwiesen, da die Konstruktion den Beweis nur unnötig verlängern und so vom eigentlichen Ziel ablenken würde.

Nimmt man das bisher simpelste Beispiel einer Glättungsfunktion (mal abgesehen von der Identität)

$$\hat{g}(w') := g^{**}(w') = \begin{cases} \underline{w}_0 + \alpha(w' - \underline{w}_0) & , w' < \underline{w}_0 \\ w' & , \underline{w}_0 \leq w' \leq \overline{w}_0 \\ \overline{w}_0 + \gamma(w' - \overline{w}_0) & , \overline{w}_0 < w' \end{cases} \quad 0 < \alpha, \gamma < 1$$

heran und macht sich Gedanken, wie der dadurch hervorgerufene Effekt sich in der Darstellung der Dichte niederschlägt, so kommt man um eine Anwendung des Transformationssatzes für Dichten nicht mehr herum. Die Ausgangssituation ist mit  $X \sim \mathcal{N}(\beta w, \alpha_0 + \alpha_1 w^2)$  identisch, gesucht ist in diesem Fall jedoch die Dichte von  $\hat{g} \circ X$  oder  $f_{\hat{g} \circ X}$ .

Gemäß der Forderung im Transformationssatz und der Konstruktion der Glättungsfunktion, zerlegt man den Definitionsbereich von  $\hat{g}$  in die drei offenen Intervalle

$$I_1 = ]-\infty, \underline{w}_0[, \quad I_2 = ]\underline{w}_0, \overline{w}_0[, \quad I_3 = ]\overline{w}_0, +\infty[.$$

Mit der Einschränkung auf eines dieser drei Intervalle, hat man es jeweils mit einer stetig differenzierbaren Bijektion zu tun und kann dementsprechend die Umkehrfunktion bestimmen.

$$\hat{g}_1 = \hat{g}|_{I_1} \Rightarrow \hat{g}_1^{-1}(y) = \frac{y - \underline{w}_0}{\alpha} + \underline{w}_0$$

$$\hat{g}_2 = \hat{g}|_{I_2} \Rightarrow \hat{g}_2^{-1}(y) = \text{Id}(y) = y$$

$$\hat{g}_3 = \hat{g}|_{I_3} \Rightarrow \hat{g}_3^{-1}(y) = \frac{y - \overline{w}_0}{\gamma} + \overline{w}_0$$

Offenbar ist auch  $\frac{\partial g_m(x)}{\partial x} \neq 0$  für  $m \in \{1, 2, 3\}$  erfüllt und

$$\left| \frac{\partial \hat{g}_1^{-1}(y)}{\partial y} \right| = \left| \frac{1}{\alpha} \right| = \frac{1}{\alpha} > 0,$$

$$\left| \frac{\partial \hat{g}_2^{-1}(y)}{\partial y} \right| = |1| = 1 > 0,$$

$$\left| \frac{\partial \hat{g}_3^{-1}(y)}{\partial y} \right| = \left| \frac{1}{\gamma} \right| = \frac{1}{\gamma} > 0.$$

Damit sind alle Voraussetzungen gegeben und der Transformationssatz liefert die Dichte

$$f_{\hat{g} \circ X}(y) = \begin{cases} f_X(\hat{g}_1^{-1}(y)) \cdot \left| \frac{\partial \hat{g}_1^{-1}(y)}{\partial y} \right| = \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((\frac{y-w_0}{\alpha} + w_0) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\alpha} \right| & \text{für } y < \underline{w_0} \\ f_X(\hat{g}_2^{-1}(y)) \cdot \left| \frac{\partial \hat{g}_2^{-1}(y)}{\partial y} \right| = \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(y - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot |1| & \text{für } \underline{w_0} < y < \overline{w_0} \\ f_X(\hat{g}_3^{-1}(y)) \cdot \left| \frac{\partial \hat{g}_3^{-1}(y)}{\partial y} \right| = \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((\frac{y-\overline{w_0}}{\gamma} + \overline{w_0}) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\gamma} \right| & \text{für } \overline{w_0} < y. \end{cases}$$

In Analogie zu (6), gilt es jetzt zu betrachten:

$$\int_{-\infty}^{+\infty} |y|^\eta f_{\hat{g} \circ X}(y) dy \quad (7)$$

Dies führt, mit den gleichen Schritten wie oben und nach Anwendung des Transformationssatzes, zu einer vergleichbaren Darstellung für (7). Die Details der Umformungsschritte und Erläuterungen zu diesem langwierigen Unterfangen, sind im Anhang (Abschnitt 4.2) zu finden. In der so gewonnenen Darstellungen für (6) und (7) wird ein direkter Vergleich möglich, da sich die Integraleinträge bis auf ihre Vorfaktoren gleichen.



$$\begin{aligned}
(7) &= \int_0^{\underline{w_0}} (\alpha \cdot (x - \underline{w_0}) + \underline{w_0})^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(x-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx & (\text{i}^{***}) \\
&+ \int_{\underline{w_0}}^{\overline{w_0}} x^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(x-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx & (\text{ii}^{***}) \\
&+ \int_{\overline{w_0}}^{+\infty} (\gamma \cdot (x - \overline{w_0}) + \overline{w_0})^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(x-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx & (\text{iii}^{***}) \\
&+ \int_0^{\underline{w_0}} (\alpha \cdot (x - \underline{w_0}) + \underline{w_0})^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((-x)-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx & (\text{iv}^{***}) \\
&+ \int_{\frac{\overline{w_0}-\underline{w_0}}{\alpha} + \underline{w_0}}^{\underline{w_0}} (\alpha \cdot (x - \underline{w_0}) + \underline{w_0})^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((-x)-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx & (\text{v}^{***}) \\
&+ \int_{\frac{\overline{w_0}-\underline{w_0}}{\alpha} + \underline{w_0}}^{+\infty} (\alpha \cdot (x - \underline{w_0}) + \underline{w_0})^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((-x)-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx & (\text{vi}^{***}) \\
&+ \int_0^{\frac{\underline{w_0}}{\alpha} - \underline{w_0}} (\alpha \cdot ((-x) - \underline{w_0}) + \underline{w_0})^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((-x)-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx & (\text{vii}^{***}) \\
&+ \int_{\frac{\underline{w_0}}{\alpha} - \underline{w_0}}^{\underline{w_0}} (\alpha \cdot ((-x) - \underline{w_0}) + \underline{w_0})^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(x-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx & (\text{viii}^{***})
\end{aligned}$$

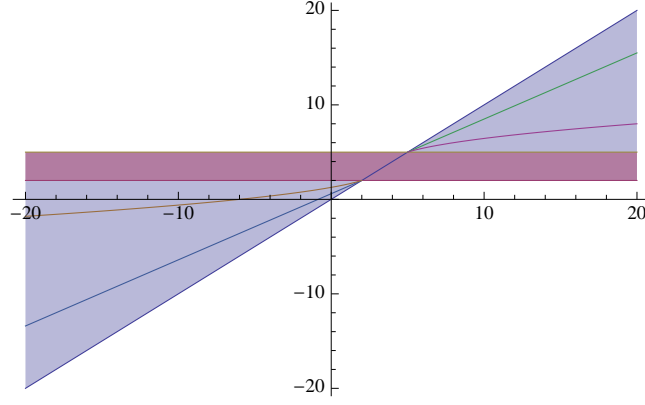
Die Begründung für die gemachten Schlussfolgerungen, findet sich wiederum im Anhang (Abschnitt 4.2). An dieser Stelle interessiert, in erster Linie, nur das Ergebnis:

$$1 + (7) \leq \left[ (1 - \delta) + \frac{\underline{w_0}^\alpha}{1 + |\underline{w}|^\eta} \right] \cdot (1 + |w|^\eta) \leq \left[ (1 - \delta) + \frac{\underline{w_0}^\alpha}{1 + |M|^\eta} \right] \cdot (1 + |w|^\eta) \leq (1 - \delta^*) h(w)$$

bei der Festlegung  $\delta^* = (\delta - \frac{\underline{w_0}^\alpha}{1 + |M|^\eta})$ . □

Jede oben als „zulässig“ erkannte Glättungsfunktion (blau schraffierter Bereich), verstärkt nur die für den Beweis verwendeten Effekte, in Hinblick auf die Verschiebung der Wahrscheinlichkeitsmasse, im Vergleich zu einer aus Geradenstücken entstandenen Glättungsfunktion.

Ganz anschaulich kann man sich diesen Sachverhalt noch mal an einer Grafik verdeutlichen



und erkennt für die Glättungsfunktion unterhalb von  $\underline{w}_0$ , bei entsprechender Wahl von  $\alpha$  und analog oberhalb von  $\overline{w}_0$ , bei entsprechender Wahl von  $\gamma$  eine Art „Majorante“ - jeweils in Form einer (Geraden-) Glättungsfunktion. Es erübrigen sich jegliche zusätzlichen Beweisschritte, indem man jede „zulässige“ Glättungsfunktion als „eingeschachtelt“ durch ein Geradenbeispiel auffasst.

Mit den bisherigen Ergebnissen, lassen sich dann auch die Unterschiede in den Parameterbereichen für den ungeglätteten und geglätteten Prozess zumindest teilweise aus der Welt schaffen. Der geglättete Prozess erfüllt nämlich die Driftbedingung, mit den selben Parameterkonstellationen, wie es auch der ungeglättete Prozess tut.

Bevor wir uns damit auseinandersetzen, in wie weit das „Glätten“ diesen Parameterbereich erweitert, zuvor noch etwas zu dem anfangs eingeführten Glättungsbeispiel:

$$g^*(w') = \begin{cases} e^{-(\underline{w}_0 - w')} + (\underline{w}_0 - 1) & , w' < \underline{w}_0, \quad \underline{w}_0 \geq 1 \\ w' & , \underline{w}_0 \leq w' \leq \overline{w}_0 \\ (\overline{w}_0 - 1) + (w' - \overline{w}_0 + 1)^\alpha & , \overline{w}_0 < w' \quad 0 < \alpha < 1. \end{cases}$$

Dieses war ja von der Praxis her motiviert und brachte nur positive Werte hervor. Auch wenn dieser Fall durch obige Überlegungen abgedeckt ist, machen die positiven Werte eine etwas schnellere Interpretation möglich und gerade deshalb, soll der Fall nicht unerwähnt bleiben.

$$\begin{aligned} g_1^* &= g^*|_{I_1} \Rightarrow g_1^{*-1}(y) = \ln(y + 1 - \underline{w}_0) + \underline{w}_0 \\ g_2^* &= g^*|_{I_2} \Rightarrow g_2^{*-1}(y) = \text{Id}(y) = y \\ g_3^* &= g^*|_{I_3} \Rightarrow g_3^{*-1}(y) = (y - \overline{w}_0 + 1)^{\frac{1}{\alpha}} + \overline{w}_0 - 1 \end{aligned}$$

Offenbar ist auch  $\frac{\partial g_m^*(x)}{\partial x} \neq 0$  für  $m \in \{1, 2, 3\}$  erfüllt und

$$\begin{aligned} \left| \frac{\partial g_1^{*-1}(y)}{\partial y} \right| &= \left| \frac{1}{y + 1 - \underline{w}_0} \right| = \frac{1}{y + 1 - \underline{w}_0} > 0 \quad \text{für } y \in ]\underline{w}_0 - 1, \underline{w}_0[ \\ \left| \frac{\partial g_2^{*-1}(y)}{\partial y} \right| &= |1| = 1 > 0 \\ \left| \frac{\partial g_3^{*-1}(y)}{\partial y} \right| &= \left| \frac{1}{\alpha} (y - \overline{w}_0 + 1)^{\frac{1}{\alpha} - 1} \right| = \frac{1}{\alpha} (y - \overline{w}_0 + 1)^{\frac{1}{\alpha} - 1} > 0 \quad \text{für } y \in ]\overline{w}_0, \infty[. \end{aligned}$$

Damit sind alle Voraussetzungen gegeben, und der Transformationssatz liefert die Dichte

$$f_{g^* \circ X}(y) = \begin{cases} 0 & \text{für } y \leq \underline{w}_0 - 1 \\ \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(\ln(y+1-\underline{w}_0) + \underline{w}_0 - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{y+1-\underline{w}_0} \right| & \text{für } \underline{w}_0 - 1 < y < \underline{w}_0 \\ \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(y-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot |1| & \text{für } \underline{w}_0 < y < \overline{w}_0 \\ \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((y-\overline{w}_0)+1)^{\frac{1}{\alpha}} + \overline{w}_0 - 1 - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\alpha} (y - \overline{w}_0 + 1)^{\frac{1-\alpha}{\alpha}} \right| & \text{für } \overline{w}_0 < y. \end{cases}$$

Im Unterschied zu den bisher behandelten Beispielen hat die Betragsfunktion hier keinen Effekt mehr, da ohnehin nur positive Werte angenommen werden. Aus

$$\int_{-\infty}^{+\infty} |y|^\eta f_{g^* \circ X}(y) dy$$

wird demnach

$$\int_{\underline{w}_0 - 1}^{+\infty} y^\eta f_{g^* \circ X}(y) dy \quad (8)$$

und die Wahrscheinlichkeitsmasse, die zuvor auf  $\mathbb{R}_-$  entfallen war, wird bereits durch die Glättungsfunktion dem Bereich  $]\underline{w}_0 - 1, \underline{w}_0[$ , zugeordnet (nicht über ganz  $\mathbb{R}_+$  verteilt).

$$\begin{aligned} (8) &= \int_{\underline{w}_0 - 1}^{\underline{w}_0} y^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(\ln(y+1-\underline{w}_0) + \underline{w}_0 - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \frac{1}{y+1-\underline{w}_0} dy \\ &+ \int_{\underline{w}_0}^{\overline{w}_0} y^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(y-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dy \\ &+ \int_{\overline{w}_0}^{\infty} y^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((y-\overline{w}_0)+1)^{\frac{1}{\alpha}} + \overline{w}_0 - 1 - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \frac{1}{\alpha} (y - \overline{w}_0 + 1)^{\frac{1-\alpha}{\alpha}} dy \end{aligned}$$

Mit der Substitution

$$\ln(y+1-\underline{w}_0) + \underline{w}_0 = t \quad \text{bzw.} \quad (y - \overline{w}_0 + 1)^{\frac{1}{\alpha}} + \overline{w}_0 - 1 = z$$

folgt dann

$$\begin{aligned}
(8) = & \int_0^{\overline{w_0}} (e^{z-\overline{w_0}} + \overline{w_0} - 1)^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(t-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dt \\
& + \int_{\overline{w_0}}^{\overline{w_0}} y^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(y-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dy \\
& + \int_{\overline{w_0}}^\infty ((z+1-\overline{w_0})^\alpha - 1 + \overline{w_0})^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(z-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dz
\end{aligned}$$

und dank  $(z+1-\overline{w_0})^\alpha - 1 + \overline{w_0} < z$  für  $z \in (\overline{w_0}, \infty)$  sind identische Schlussfolgerungen wie oben möglich. Die Situation ist aber sehr viel eindeutiger als zuvor, da die gesamte Wahrscheinlichkeitsmasse von  $\mathbb{R}_-$  nicht mehr auf  $\mathbb{R}_+$  umgelegt wird und somit auch die Auswirkungen auf den Erwartungswert klarer sind.

## 2.3 Was Glätten tatsächlich bringt

### 2.3.1 Parameterbereicherweiterung dank Glättung mit Geradenstücken

Die obigen Beweisschritte legen noch einmal sehr deutlich dar, wie die unterschiedlichen Glättungsfunktionen zu verknüpfen sind und stellen die Grundidee der Glättung auf ein solides Fundament. Um zu erkennen, in welchem Ausmaß das Glätten den Bereich für geometrische Ergodizität erweitert, sind sie jedoch wenig hilfreich. Dies liegt zum Teil auch daran, dass sich die von Borkovec und Klüppenberg [19] an das AR(1)-ARCH(1)-Modell gestellte Hauptbedingung

$$\mathbb{E}(\log |\beta + \sqrt{\alpha_1} \varepsilon|) < 0$$

nicht unbedingt schnell und natürlich herleiten lässt. Man denke nur an die Plausibilitätsüberlegungen<sup>4</sup> und die Tatsache, dass ein konkreter Parameterbereich nur mittels Iterationsverfahren angegeben werden kann. Will man jedoch den „Wert“ des Glättungsansatzes irgendwie beschreiben, so sind scharf abgrenzbare Werte von großem Vorteil.

Die Ergebnisse von BuK [19] beruhen zum Teil auf bzw. sind inspiriert von Maerker [23] und auch wenn die dortigen Ergebnisse nicht so umfassend sind, bieten sie einen viel schnelleren Einblick in ihre Entstehung. Die Darstellung orientiert sich hierbei an der Driftbedingung, wie sie in Meyn und Tweedie [34] eingeführt wird.

**Satz 2.9.** (*geometrische Ergodizität AR(1)-ARCH(1) - Lemma 2.1.2 [23]*)

Für die Testfunktion  $h(w) = 1 + w^2$  ergibt sich geometrische Ergodizität im AR(1)-ARCH(1)-Fall  $(W_t, t \in \mathbb{N})$  unter der Bedingung:

$$\beta^2 + \alpha_1 < 1$$

---

<sup>4</sup>vgl.  $\beta, \alpha_1$  Zusammenhang im Anhang

**Beweis zu Satz 2.9.** (vgl. Proof zu Lemma 2.1.2 - [23])

$$\begin{aligned}
& \frac{\mathbb{E}[h(W_t) \mid W_{t-1} = w] - h(w)}{h(w)} \\
&= \frac{\mathbb{E}[1 + (\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t)^2] - (1 + w^2)}{1 + w^2} \\
&= \frac{\mathbb{E}[1 + \beta^2 w^2 + 2\beta w \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t + (\alpha_0 + \alpha_1 w^2) \varepsilon_t^2] - 1 - w^2}{1 + w^2} \\
&= \frac{\beta^2 w^2 + (\alpha_0 + \alpha_1 w^2) - w^2}{1 + w^2} \quad \text{da } \mathbb{E}(\varepsilon_t) = 0 \text{ und } \mathbb{E}(\varepsilon_t^2) = 1 \\
&= \frac{\beta^2 w^2 + \alpha_1 w^2}{1 + w^2} + \frac{\alpha_0}{1 + w^2} - \frac{w^2}{1 + w^2} \\
&< \frac{\beta^2 w^2 + \alpha_1 w^2}{w^2} + \frac{\alpha_0}{1 + w^2} - \frac{w^2}{1 + w^2} \\
&\leq -\delta
\end{aligned}$$

Für  $|w| > M$ , sofern die Bedingung

$$\limsup_{|w| \rightarrow \infty} \frac{\beta^2 w^2 + \alpha_1 w^2}{w^2} < 1$$

oder einfach

$$\boxed{\beta^2 + \alpha_1 < 1} \tag{9}$$

erfüllt ist. □

Mit (9) genügt eine hinreichend große Schranke  $M$  und die Tatsache, dass

$$\lim_{w \rightarrow \infty} \frac{\alpha_0}{1 + w^2} \searrow 0, \quad \lim_{w \rightarrow \infty} \frac{w^2}{1 + w^2} \nearrow 1,$$

um die Existenz von  $\delta > 0$  zu sichern. Diese Schranke  $M$  aus dem Beweis, liefert auch gleich die notwendige Menge  $C = \{w \in \mathbb{R} : |w| \leq M\}$  für die Driftbedingung.

Wie oben bereits an anderer Stelle, soll auch hier auf die durch Geradenstücke hervorgerufene Glättung zurückgegriffen werden, um die Beziehung zwischen geglätteten und ungeglätteten Fall herzustellen.

**Satz 2.10.** (geometrische Ergodizität Smooth-AR(1)-ARCH(1))

Für die Testfunktion  $h(w) = 1 + w^2$  ergibt sich geometrische Ergodizität im Smooth-AR(1)-ARCH(1)-Fall ( $\tilde{W}_t, t \in \mathbb{N}$ ) - entstanden durch die Geradenglättung, unter der Bedingung:

$$\zeta^2 \beta^2 + \zeta^2 \alpha_1 < 1$$

**Beweis zu Satz 2.10.**

Erinnert man sich an das Konstruktionsprinzip, so ergibt sich für den geglätteten Fall  $(\tilde{W}_t, t \in \mathbb{N})$  zunächst:

$$\begin{aligned} & \frac{\mathbb{E}[h(\tilde{W}_t) \mid \tilde{W}_{t-1} = w] - h(w)}{h(w)} \\ &= \frac{\mathbb{E} \left[ 1 + \left( \left( \beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t \right) \mathbf{1}_K + \left( \underline{w}_0 + \alpha(\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t - \underline{w}_0) \right) \mathbf{1}_L + \left( \overline{w}_0 + \gamma(\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t - \overline{w}_0) \right) \mathbf{1}_R \right)^2 \right] - (1+w^2)}{1+w^2} \end{aligned}$$

$\mathbf{1}_L, \mathbf{1}_K, \mathbf{1}_R =$  Indikatorfunktion links, rechts oder genau im Kompaktum (vgl. oben).

Löst man den quadratischen Ausdruck, verschwinden wegen der Indikatorfunktion die gemischten Terme  $(\mathbf{1}_A \cdot \mathbf{1}_B = \mathbf{1}_{A \cap B})$ .

$$= \frac{\mathbb{E} \left[ 1 + \left( \beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t \right) \mathbf{1}_K^2 + \left( \underline{w}_0 + \alpha(\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t - \underline{w}_0) \right) \mathbf{1}_L^2 + \left( \overline{w}_0 + \gamma(\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t - \overline{w}_0) \right) \mathbf{1}_R^2 \right] - (1+w^2)}{1+w^2}$$

Im Vergleich zu dem obigen Beweis des ungeglätteten Falls, muss man sich erst einmal nur vor Augen halten, dass die Werte innerhalb des Kompaktums beschränkt sind. Sie ließen sich also, in Analogie zu oben, mit einer an  $\max\{|\underline{w}_0|, |\overline{w}_0|\}$  angelehnten Konstanten behandeln.

Nur die beiden verbliebenen Ausdrücke, links und rechts des Kompaktums, können beliebig „große“ Werte generieren. Hierbei unterscheiden diese sich konstruktionsgemäß nur in den Komponenten  $(\underline{w}_0, \alpha), (\overline{w}_0, \gamma)$ . Diese Ähnlichkeit und die quadratischen Terme erleichtern die weiteren Schlussfolgerungen. Anstatt also zwei disjunkte Fälle zu behandeln, wollen wir diese zusammenfassen, indem  $\underline{w}$  als Platzhalter für die möglichen Konstanten  $\underline{w}_0, \overline{w}_0$  und  $\zeta = \max\{\alpha, \gamma\}$  genutzt werden.

$$\begin{aligned} & \leq \frac{\mathbb{E} \left[ 1 + \left( \max\{|\underline{w}_0|, |\overline{w}_0|\} \right)^2 + \left( \underline{w} + \zeta(\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t - \underline{w}) \right) \mathbf{1}_{L,R} \right]^2 - (1+w^2)}{1+w^2} \\ &= \frac{\mathbb{E} \left[ 1 + \left( \max\{|\underline{w}_0|, |\overline{w}_0|\} \right)^2 + \left( \underline{w}^2 + 2\underline{w}\zeta(\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t - \underline{w}) + \zeta^2 \left( (\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t)^2 - 2\underline{w}(\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t) + \underline{w}^2 \right) \right) \mathbf{1}_{L,R} \right] - (1+w^2)}{1+w^2} \\ & \leq \frac{\mathbb{E} \left[ 1 + \left( \max\{|\underline{w}_0|, |\overline{w}_0|\} \right)^2 + \left( \underline{w}^2 + 2\underline{w}\zeta(\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t - \underline{w}) + \zeta^2 \left( (\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t)^2 - 2\underline{w}(\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t) + \underline{w}^2 \right) \right) \right] - (1+w^2)}{1+w^2} \\ &= \frac{(\max\{|\underline{w}_0|, |\overline{w}_0|\})^2 + \underline{w}^2 + 2\underline{w}\zeta\beta w - 2\underline{w}^2\zeta + \zeta^2\beta^2 w^2 + \zeta^2(\alpha_0 + \alpha_1 w^2) - \zeta^2 2\underline{w}\beta w + \zeta^2 \underline{w}^2 - w^2}{1+w^2} \\ & \leq \frac{\zeta^2\beta^2 w^2 + \zeta^2\alpha_1 w^2}{w^2} + \frac{(\max\{|\underline{w}_0|, |\overline{w}_0|\})^2 + \zeta^2\alpha_0 - \zeta^2 2\underline{w}\beta w + \zeta^2 \underline{w}^2 + \underline{w}^2 + 2\underline{w}\zeta\beta w - 2\underline{w}^2\zeta}{1+w^2} - \frac{w^2}{1+w^2} \\ & \leq -\delta \end{aligned}$$

für  $|w| > \tilde{M}$ , sofern die Bedingung

$$\limsup_{|w| \rightarrow \infty} \frac{\zeta^2\beta^2 w^2 + \zeta^2\alpha_1 w^2}{w^2} < 1$$

oder einfach

$$\boxed{\zeta^2 \beta^2 + \zeta^2 \alpha_1 < 1} \quad (10)$$

erfüllt ist.  $\square$

Die Argumentation zur Existenz eines solchen  $\delta > 0$  entspricht genau der von oben. Insbesondere an der Darstellung in der letzten Zeile, zeigt sich  $\tilde{M} \geq \max\{|\underline{w}_0|, |\overline{w}_0|\}$ . Welcher der beiden Werte betragsmässig der größere ist und zum Glättungsansatz von oben oder von unten gehört, ist jedoch nicht entscheidend. Auch der Umstand, dass nur der größere der beiden Glättungsfaktoren (Stellschrauben) für das Hauptkriterium von Bedeutung ist, kann an dieser Stelle wunderbar abgelesen werden.

Viel wichtiger erscheint jedoch die Tatsache, dass sich die Hauptbedingung, in Sachen Interpretationsfähigkeit, im Vergleich zu BuK [19] deutlich hervorhebt. Der Umfang in dem die Glättung eine Vergrößerung des Parameterbereichs bedeutet, wird sofort klar, wenn man sich den Wertebereich für die Glättungsparameter vor Augen hält. Da  $\zeta$  zwischen 0 und 1 liegt, gilt das natürlich auch für  $\zeta^2$  und die hinreichenden Bedingungen an  $\beta$  bzw.  $\alpha_1$  können abgeschwächt werden. So kann man beispielsweise einen Wert  $\beta > 1$  zulassen, wenn dieser durch einen entsprechend kleinen Wert für  $\zeta$  ausgeglichen wird. In Hinblick auf die Grundidee des Glättens, bedeutet dies, dass man wie erhofft die Forderungen an den Prozess im Inneren des Kompaktums zurückschrauben kann.

In diesem Beispiel kann man den Zulässigkeitsbereich für  $\beta$  und  $\alpha_1$  sogar ganz konkret angeben, in Abhängigkeit von der Wahl für  $\zeta$ :

$$\boxed{\beta^2 + \alpha_1 < \frac{1}{\zeta^2}}.$$

Letztendlich liefert obiges Ergebnis die Hauptbedingung (iii) für Theorem 2 - Feigin und Tweedie [32] und damit nicht nur die geometrische Ergodizität, sondern auch die Existenz des zweiten Moments für die stationäre Verteilung.

Um sich von dieser Aussage zu überzeugen, müssen nur noch die Punkte (i) und (ii) überprüft werden. Die Bedingung (i) ergibt sich hierbei sofort für  $h(w) = 1 + w^2$  und das Kompaktum  $A = [-\tilde{M}, \tilde{M}]$ . Die Funktion  $h$  ist nicht nur  $\geq 1$ , sondern auch stetig und somit beschränkt über einem Kompaktum. Es bleibt zu überprüfen, ob auch der Punkt (ii):

$$\sup \mathbb{E}[h(\tilde{W}_t) \mid \tilde{W}_{t-1} = w] < \infty,$$

erfüllt ist für  $w \in A$ . Dazu muss im obigen Beweis nur die drittletzte Zeile nach  $\mathbb{E}[h(\tilde{W}_t) \mid \tilde{W}_{t-1} = w]$  aufgelöst werden. Ersetzt man alle negativen, oder vom Vorzeichen von  $w$  abhängenden Ausdrücke durch ihre Beträge, so ergibt sich folgende obere Schranke:

$$1 + (\max\{|\underline{w}_0|, |\overline{w}_0|\})^2 + \underline{w}^2 + 2\underline{w}\zeta\beta|w| + 2\underline{w}^2\zeta + \zeta^2\beta^2w^2 + \zeta^2(\alpha_0 + \alpha_1w^2) + \zeta^22\underline{w}\beta|w| + \zeta^2\underline{w}^2$$

Alle Einträge bis auf  $|w|$  bzw.  $w^2$  sind konstant und da der Wertebereich für  $w$  durch das Kompaktum  $A$  beschränkt ist, bleibt auch der Gesamtausdruck beschränkt.

Im Zusammenhang mit obiger Driftbedingung werden stets Eigenschaften wie  $\phi$ -Irreduzibilität, Aperiodizität und die sogenannten „small“-sets benötigt (vgl. z.B. Theorem 1 - [32]). Bei vorausgesetzter Normalverteilung für die Störterme, ergibt sich mit dem Lebesgue-Maß die geforderte  $\lambda$ -Irreduzibilität. Aperiodizität und die Eigenschaft einer Feller-Kette, ergeben sich konstruktionsbedingt (vgl. [13], [32]) und werden durch die Glättungsfunktion nicht beeinflusst. Hierbei kommt insbesondere die Stetigkeitsbedingung an die Glättungsfunktionen zum tragen - vgl. Abschnitt 4.4.

Zusammenfassend ergibt sich durch den Glättungsansatz eine Erweiterung des zulässigen Parameterbereichs, für geometrische Ergodizität und auch für die Existenz des zweiten Moments der stationären Verteilung. Kennt man also gewisse Parameter der zugrundeliegenden Situation, wie etwa die oben angesprochenen „kleinsten“ oder „größten“ Preise, so kann man den zulässigen Parameterbereich vergrößern und damit womöglich eine bessere Anpassung an gegebenes Datenmaterial erreichen. Unter Umständen ist es möglich, mit einem verhältnismässig einfachen Modell und einem Glättungsansatz zu einer guten Anpassung zu kommen, die sonst nur über einen Modellwechsel möglich gewesen wäre (vgl. empirische Studie).

Die Glättungsfunktion mittels Geradenstücken ist ja hervorgegangen aus der Überlegung die anfangs eingeführten asymmetrischen Glättungsfunktionen, wie beispielsweise

$$g(w') = \begin{cases} \underline{w}_0 e^{-(\underline{w}_0 - w')} & , w' < \underline{w}_0 \\ w' & , \underline{w}_0 \leq w' \leq \overline{w}_0 \\ \overline{w}_0 + \gamma(w' - \overline{w}_0) & , \overline{w}_0 < w' \end{cases}$$

$$0 < \gamma < 1$$

auf ihre Minimalbedingungen hin zu untersuchen. Auch wenn der asymmetrische Aufbau komplizierter wirkt, verhält es sich in Hinblick auf den Beweis genau anders herum. Im Beweis zu Satz 2.10 kam die Symmetrie dahingehend zum tragen, dass das Verhalten links und rechts des Kompaktums in einem Fall zusammengefasst werden konnte. Die Asymmetrie in der obigen Glättungsfunktion verhindert zwar eine solche Vorgehensweise, macht aber durch die besondere Konstruktion (Exponentialfunktion) eine analoge Strategie möglich. So kann man die Betrachtung des linken Zweigs durch eine Konstante ersetzen (hier zum Beispiel  $\underline{w}_0$ ) und sich voll und ganz auf den rechten Zweig konzentrieren. Demzufolge ist auch die Festlegung  $\zeta = \max\{\alpha, \gamma\}$  unnötig und das konkrete Ergebnis in Satz 2.10 würde sich verändern zu:

$$\boxed{\gamma^2 \beta^2 + \gamma^2 \alpha_1 < 1} \tag{11}$$

Der zulässige Bereich sich wiederum erweitern zu:

$$\boxed{\beta^2 + \alpha_1 < \frac{1}{\zeta^2}}$$

Im Konsens bedeutet dies, dass nur der Glättungsparameter von oben einschränkend wirkt und der Beweis sich sogar minimal vereinfacht/verkürzt.



### 2.3.2 Polynomiale Glättung

Die bisherigen Abwandlungen der Glättungsidee hatten auch immer die Zielsetzung, die notwendigen Mechanismen des Glättens zu erkennen und minimale Bedingungen aufzustellen. Dies führte letztendlich zum Glättungsansatz mittels Geradenstücken. Weitere Beweise für die verschiedenen Glättungsansätze sind damit eigentlich hinfällig, sofern man sich diese immer als eingeschachtelt durch ein Geradenbeispiel vorstellt. Mit dem obigen Beweis ist man nun aber in der Situation, den Glättungseffekt tatsächlich in Form einer Parameterbereicherweiterung ablesen zu können und vor diesem Hintergrund erscheint eine analoge Untersuchung der „stärkeren“ Glättungsfunktionen äußerst sinnvoll.

Um den obigen Beweis so weit wie möglich adaptieren zu können, wird folgende Glättungsfunktion untersucht:

$$g^{4*}(w') = \begin{cases} (\underline{w}_0 + 1) - (\underline{w}_0 + 1 - w')^\alpha & , w' < \underline{w}_0 \\ w' & , \underline{w}_0 \leq w' \leq \overline{w}_0 \\ (\overline{w}_0 - 1) + (w' - \overline{w}_0 + 1)^\gamma & , \overline{w}_0 < w' \end{cases} \quad 0 < \alpha, \gamma < 1.$$

Man vergleiche dies, insbesondere mit der Konstruktion von  $g^{**}(\cdot)$ . Der Glättungseffekt ist nun oberhalb und unterhalb des Kompaktums identisch, bis auf die möglicherweise unterschiedlichen Parameter  $\alpha, \gamma$ . Dieses Konstrukt lässt sich mit dem Geradenansatz besser in Einklang bringen, da es in beiden Fällen keine Einschränkung auf den positiven Bereich gibt, aber die polynomiale Glättung ist, insbesondere für Werte weit außerhalb des Kompaktums, wesentlich stärker. Zu welcher Veränderung des Parameterbereichs diese neue Glättungsfunktion führt, soll der direkte Vergleich der Beweise zeigen.

Der Einstieg in den Beweis erfolgt zunächst völlig analog zu oben  $((\tilde{W}_t, t \in \mathbb{N})$  durch  $g^{4*}$ ):

$$\begin{aligned} & \frac{\mathbb{E}[h(\tilde{W}_t) \mid \tilde{W}_{t-1} = w] - h(w)}{h(w)} \\ = & \frac{\mathbb{E}\left[1 + \left((\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t)\right)_{\mathbb{1}_K} + \left((\underline{w}_0 + 1) - [\underline{w}_0 + 1 - (\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t)]^\alpha\right)_{\mathbb{1}_L} + \left((\overline{w}_0 - 1) + [\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t - \overline{w}_0 + 1]^\gamma\right)_{\mathbb{1}_R}\right]^2}{1 + w^2} - (1 + w^2) \end{aligned}$$

$\mathbb{1}_L, \mathbb{1}_K, \mathbb{1}_R$  = Indikatorfunktion links, rechts oder genau im Kompaktum (vgl. oben).

Löst man den quadratischen Ausdruck, verschwinden wegen der Indikatorfunktion die gemischten Terme ( $\mathbb{1}_A \cdot \mathbb{1}_B = \mathbb{1}_{A \cap B}$ ).

$$= \frac{\mathbb{E}\left[1 + \left(\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t\right)_{\mathbb{1}_K}^2 + \left((\underline{w}_0 + 1) - [\underline{w}_0 + 1 - (\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t)]^\alpha\right)_{\mathbb{1}_L}^2 + \left((\overline{w}_0 - 1) + [\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t - \overline{w}_0 + 1]^\gamma\right)_{\mathbb{1}_R}^2\right]}{1 + w^2} - (1 + w^2)$$

Die Werte im Kompaktum lassen sich wiederum mittels einer Konstanten  $\max\{|\underline{w}_0|, |\overline{w}_0|\}$  abschätzen, die verbliebenen Terme zusammenfassen und abschätzen, durch die Festlegung:  $\overline{w}$  als Platzhalter für  $\{(\underline{w}_0 + 1), (\overline{w}_0 - 1)\}$ , sowie  $\zeta = \max\{\alpha, \gamma\}$ .

$$\leq \frac{\mathbb{E}\left[1 + (\max\{|\underline{w}_0|, |\overline{w}_0|\})^2 + \left(\overline{w} \pm [\pm(\overline{w} - (\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t))]^\zeta\right)_{\mathbb{1}_{L,R}}^2\right]}{1 + w^2} - (1 + w^2)$$

$$\begin{aligned}
&= \frac{\mathbb{E} \left[ 1 + \left( \max\{|\underline{w}_0|, |\overline{w}_0|\} \right)^2 + \overline{w}^2 \pm 2\overline{w} \left[ \pm \left( \overline{w} - (\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t) \right) \right]_{\mathbb{1}_{L,R}}^{\zeta} + \left( \overline{w} - (\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t) \right)_{\mathbb{1}_{L,R}}^{2\zeta} \right] - (1+w^2)}{1+w^2} \\
&\leq \frac{\mathbb{E} \left[ 1 + \left( \max\{|\underline{w}_0|, |\overline{w}_0|\} \right)^2 + \overline{w}^2 \pm 2\overline{w} \left( \overline{w} - (\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t) \right)_{\mathbb{1}_{L,R}} + \left( \overline{w} - (\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t) \right)_{\mathbb{1}_{L,R}}^{2\zeta} \right] - (1+w^2)}{1+w^2}
\end{aligned}$$

da  $\pm \left( \overline{w} - (\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t) \right) > 1$  und  $0 < \zeta < 1$

$$\leq \frac{1 + \left( \max\{|\underline{w}_0|, |\overline{w}_0|\} \right)^2 + \overline{w}^2 \pm 2\beta w \overline{w} + \mathbb{E} \left[ \left( \overline{w} - (\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t) \right)^{2\zeta} \right] - (1+w^2)}{1+w^2}$$

wobei aus  $\varepsilon_t \sim \mathcal{N}(0, 1)$  folgt  $(\overline{w} - (\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon_t)) \sim \mathcal{N}(\mu, \sigma^2)$   
mit  $\mu = \overline{w} - \beta w$  und  $\sigma^2 = \alpha_0 + \alpha_1 w^2$ .

Es gilt  $\mathbb{E} \left[ \left( \mathcal{N}(\mu, \sigma^2) \right)^{2\zeta} \right] \leq \left[ \mathbb{E} \left( \mathcal{N}(\mu, \sigma^2) \right)^2 \right]^{\zeta}$  nach der Jensenschen Ungleichung, da  $(\mathcal{N}(\mu, \sigma^2))^2 \geq 0$  und  $x^{\zeta}$  eine konkave Funktion für  $x \in \mathbb{R}_{\geq 0}$ . Das zweite (nicht zentrierte) Moment der Normalverteilung ist bekanntlich  $\mu^2 + \sigma^2$  und somit:

$$\begin{aligned}
&\leq \frac{1 + \left( \max\{|\underline{w}_0|, |\overline{w}_0|\} \right)^2 + \overline{w}^2 \pm 2\beta w \overline{w} + \left( (\overline{w} - \beta w)^2 + \alpha_0 + \alpha_1 w^2 \right)^{\zeta} - (1+w^2)}{1+w^2} \\
&= \frac{\left( \max\{|\underline{w}_0|, |\overline{w}_0|\} \right)^2 + \overline{w}^2 \pm 2\beta w \overline{w} + \left( \overline{w}^2 - 2\beta w \overline{w} + \beta^2 w^2 + \alpha_0 + \alpha_1 w^2 \right)^{\zeta} - w^2}{1+w^2} \\
&\leq \frac{\left( \overline{w}^2 - 2\beta w \overline{w} + \beta^2 w^2 + \alpha_0 + \alpha_1 w^2 \right)^{\zeta}}{(w^2)^{\zeta}} + \frac{\left( \max\{|\underline{w}_0|, |\overline{w}_0|\} \right)^2 + \overline{w}^2 \pm 2\beta w \overline{w}}{1+w^2} - \frac{w^2}{1+w^2} \\
&= \left( \frac{\overline{w}^2 - 2\beta w \overline{w} + \beta^2 w^2 + \alpha_0 + \alpha_1 w^2}{w^2} \right)^{\zeta} + \frac{\left( \max\{|\underline{w}_0|, |\overline{w}_0|\} \right)^2 + \overline{w}^2 \pm 2\beta w \overline{w}}{1+w^2} - \frac{w^2}{1+w^2} \\
&\leq -\delta
\end{aligned}$$

für  $|w| > \tilde{M}$ , sofern die Bedingung

$$\limsup_{|w| \rightarrow \infty} \left( \frac{\overline{w}^2 - 2\beta w \overline{w} + \beta^2 w^2 + \alpha_0 + \alpha_1 w^2}{w^2} \right)^{\zeta} < 1$$

oder einfach

$$\boxed{(\beta^2 + \alpha_1)^{\zeta} < 1}$$

erfüllt ist.

Da letztendlich sehr großzügig mit der Jensenschen Ungleichung abgeschätzt wurde, ist an diesem kompakten Ausdruck keine Erweiterung des Parameterbereichs erkennbar.

Um auf die Anwendung der Jensenschen Ungleichung verzichten zu können, müsste man für  $[Y \sim \mathcal{N}(\mu, \sigma^2)]$ , den Ausdruck  $\mathbb{E}[Y^{2\zeta}]$  direkt berechnen oder zumindest handlich abschätzen. Mit  $0 < 2\zeta < 2$  steht man aber vor dem Problem der fraktionalen Momente (vgl. z.B. [12]) und ein geschlossener Ausdruck, wie zuvor, scheint unerreichbar. Selbst eine auf den ersten Blick

lohnenswert erscheinende Anwendung der Error-Funktion<sup>5</sup> führt nur zu kaum überschaubaren Abschätzungen, ohne klar erkennbaren Effekt oder Mehrwert.

Nichts desto trotz, möchte man ja den offenbar wesentlich stärkeren Glättungseffekt irgendwie greifbar machen und so eine Brücke schlagen, zwischen den auf den ersten Blick vielleicht gar nicht so unterschiedlich wirkenden Glättungsansätzen.

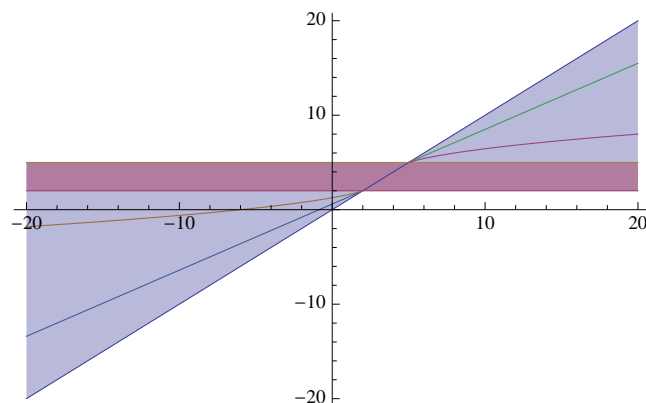
### **Satz 2.11. Polynomiale Glättung und Driftbedingung**

*Der polynomiale Glättungsansatz zeichnet sich durch eine fallende Steigung, bei zunehmender Entfernung vom unberührten Kompaktum aus. Dies schlägt sich in der Driftbedingung dahingehend nieder, dass die Parameterbereicherweiterung nur abhängig ist von der Wahl der Ausnahmemenge  $M$ . Je größer  $M$  gewählt wird, umso weiter kann der Bereich für zulässige Parameter gewählt werden.*

#### **Beweis zu Satz 2.11.**

Das folgende ist weniger als stringenter Beweis, denn mehr als Veranschaulichung zu verstehen.

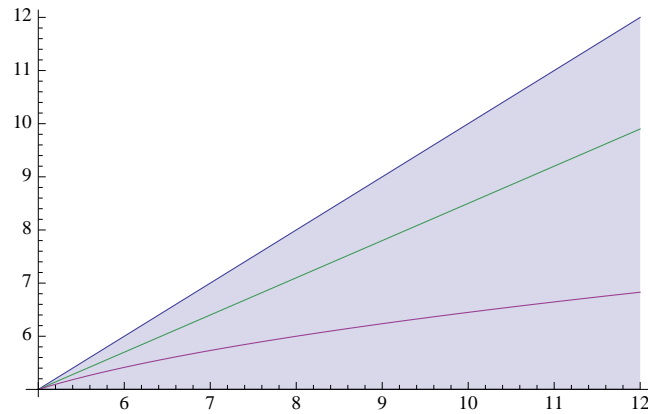
Wie einleitend schon erwähnt, kann man sich die polynomiale Glättung immer durch ein Geradenbeispiel eingeschachtelt vorstellen und hat auf diese Weise eine „Messlatte“, an der man sich orientieren kann.



Der blau schraffierte Bereich unter bzw. über der Winkelhalbierenden (Identität), war ja als der Zulässige für mögliche Glättungsfunktionen erkannt worden. Innerhalb dieses Bereichs bewegt sich auch immer die Kurve der polynomialen Glättungsfunktion.

---

<sup>5</sup>vgl. Anhang

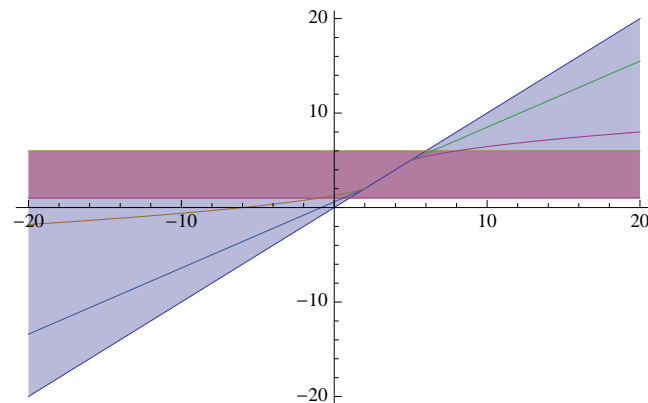


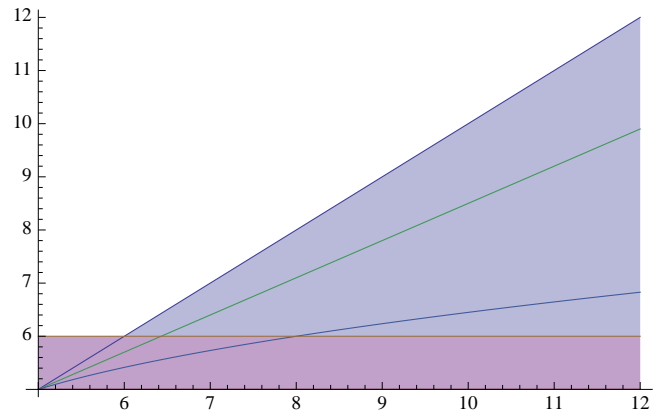
Eine Gerade zwischen Identität und polynomialer Glättungsfunktion ist dann die bereits erwähnte „Messlatte“, an der man seine Argumentation ausrichten kann und im obigen Ausschnitt der Gesamtgrafik ist diese grün markiert. Der Winkel oder die Steigung dieser Geraden ist so angelegt, dass die polynomiale Kurve stehen unterhalb bleibt - s.u..

Der Mehrwert des polynomialen Ansatzes, erschließt sich nun bei der Betrachtung der Driftbedingung außerhalb eines gewissen Kompaktums - vgl. Schlussfolgerung im Geradenbeispiel oben.

Der rot schraffierte Bereich war in unserer Betrachtung jener Part, der von der Glättung ausgenommen bleibt. Für die Beweisführung bestimmt man einen weiteren Bereich (vgl. oben:  $|w| > \tilde{M}$ ) und erst das Verhalten außerhalb dieses neuen Bereichs ist entscheidend für die Driftbedingung.

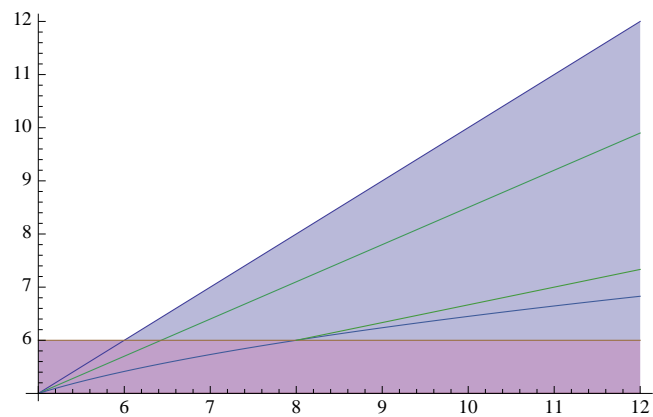
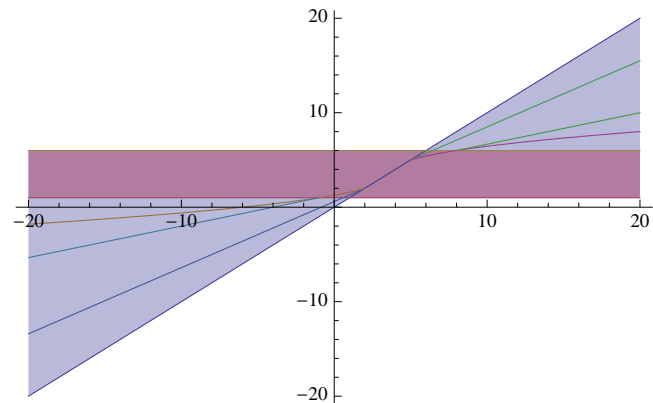
Anhand der Zeichnung macht man sich dies, durch einen breiteren roten Streifen, deutlich.





Schaut man sich zunächst nur das Geradenbeispiel an, so ist zwar der interessante Bereich weiter nach außen gerückt, die Ausgangsgerade ist aber weiterhin die begrenzende Kraft. Der Parameter der Geraden oder besser gesagt ihre Steigung ändert sich nicht und so bleibt ihr Effekt der gleiche, unabhängig davon wie weit man den roten Streifen aufbläht. Diese Tatsache spiegelt sich in der Hauptbedingung des oberen Beweises wider:  $\zeta^2\beta^2 + \zeta^2\alpha_1 < 1$ . So ist die Erweiterung des Parameterbereichs durch die Steigung (bzw. das Maximum der beiden Steigungen) der Glättungsgeraden bestimmt. Je geringer die Steigung der Geraden, umso größer der zulässige Parameterbereich für  $\beta$  und  $\alpha_1$ .

Der Unterschied im polynomialen Ansatz ist nun der, dass sich das Verhalten sehr wohl ändert, wenn der interessante Bereich weiter nach außen rückt:



Die „Messlatte“ mit der man in die Betrachtung gestartet ist, bleibt natürlich auch weiterhin gültig. In diesem Fall bietet es sich jedoch an, eine neue Gerade mit einer geringeren Steigung zum Vergleichselement zu machen. Diese geht nun von jenem Punkt aus, in dem die Glättungsfunktion den roten Bereich hinter sich lässt. So wie der rote Bereich sich zunächst durch die Grenzen  $\underline{w}_0, \overline{w}_0$  kennzeichnen ließ, kann man diese neuen Grenzen in Analogie zu der obigen Beweisführung und Notation mit  $\underline{M}_0, \overline{M}_0$  kenntlich machen. Der polynomiale Glättungsansatz lässt sich auf diese Art und Weise durch eine Anpassung der beschränkenden Geraden immer weiter annähern. Ganz konkret heißt das, anhand der oberen Grafik, dass sich die polynomiale Glättung für Werte größer als  $8 = \overline{M}_0$  durch eine Gerade mit geringerer Steigung abschätzen lässt.

Nachdem man sich mit der Idee hinreichend vertraut gemacht hat, muss man diese nur noch entsprechend formalisieren.

Im Geradenbeispiel sind bekanntermaßen nur die Ausdrücke  $\alpha \cdot (w' - \underline{w}_0)$  und  $\gamma \cdot (w' - \overline{w}_0)$  für die Steigung der Glättungsgeraden entscheidend und die jeweiligen Werte ergeben sich als Ableitung dieser Ausdrücke nach  $w'$ , also  $\alpha$  bzw.  $\gamma$ .

Im polynomialen Ansatz sind die entsprechenden Ausdrücke  $-(\underline{w}_0 + 1 - w')^\alpha$  bzw.  $(w' - \overline{w}_0 + 1)^\gamma$  und die Ableitungen:

$$\alpha \cdot \frac{1}{(\underline{w}_0 + 1 - w')^{1-\alpha}} \quad \text{bzw.} \quad \gamma \cdot \frac{1}{(w' - \overline{w}_0 + 1)^{1-\gamma}}.$$

An den Stellen  $\underline{w}_0, \overline{w}_0$  werden die Nenner der beiden Ausdrücke zu 1 und somit wiederum  $\alpha, \gamma$ .

Für  $\underline{w}_0 \neq w' \neq \overline{w}_0$  sind die Nenner jeweils  $> 1$  (beachte:  $0 < 1 - \alpha, 1 - \gamma < 1$ ), die Gesamtausdrücke dementsprechend kleiner als  $\alpha$  bzw.  $\gamma$ .

Eine Grenzwertbetrachtung  $\lim_{|w'| \rightarrow \infty}$ , erklärt die kontinuierlich fallende Steigung, je weiter man sich von dem roten Bereich entfernt.

Für die obigen Grafiken bzw. Umsetzung der Idee bedeutet dies folgendes: An den Grenzen  $\underline{w}_0, \overline{w}_0$  kann man eine polynomiale Glättung mit den Parametern  $\alpha, \gamma$  durch eine Geradenglättung mit eben diesen Parametern abschätzen - da sie genau von dieser Geraden noch überdeckt wird (vgl. „Messlatte“ oben). Mit der Erweiterung des roten Streifens, kann man eine neue Gerade als schärfere Abschätzung ansetzen und zwar mit der Steigung :  $\alpha \cdot \frac{1}{(\underline{w}_0 + 1 - \underline{M})^{1-\alpha}}$  bzw.  $\gamma \cdot \frac{1}{(\overline{M} - \overline{w}_0 + 1)^{1-\gamma}}$ .

Diesen Zusammenhang kann man sich (hier für die Blickrichtung  $+\infty$ ) auch etwas knapper notieren:

$$\forall \zeta > 0 \exists \overline{M}_0 : \quad \forall w \geq \overline{M}_0 : \quad (w - \overline{w})^\alpha \leq \zeta (w - \overline{w})$$

Man liest an dieser Formel oder den obigen Ausführungen einfach die positive, aber fallende Steigung der polynomialen Glättung ab, die sich ins unendliche fortsetzt und somit letztendlich auf ein Abschneiden hinausläuft. Sofern man den Bereich bzw.  $|w|$  nur groß genug wählt, kann

eine Gerade mit beliebig kleiner Steigung angepasst werden, um den weiteren Verlauf abzuschätzen.

Wenn eine immer kleiner werdende Steigung, einen immer größeren Zulässigkeitsbereich bedeutet, so ist der polynomiale Ansatz, so stark, dass er unabhängig von der Wahl der Parameter  $\alpha, \gamma$  immer zu der gleichen Erweiterung des Parameterbereichs führt.

Im polynomialen Fall hat man also die Situation vor sich, wie Sie im Geradenbeispiel unter der Prämisse  $\lim_{\alpha, \gamma \rightarrow 0}$  entstünde. Das augenfälligste an diesem Umstand ist die Tatsache, dass sich die Situation im polynomialen Fall völlig unabhängig von der Wahl für  $\alpha, \gamma$  so darstellt. Der polynomiale Ansatz übertrifft auf lange Sicht jedes Geradenbeispiel und hat zur Folge, dass die Parameter  $\beta, \alpha_1$  völlig frei wählbar sind, bei jeder möglichen Wahl von  $\alpha, \gamma$ .  $\square$

## 2.4 Existenz von Momenten der stationären Verteilung

Für vielfältige Anwendungen, wie zum Beispiel auch die weitere Untersuchung der Tails, ist das Wissen um die Existenz von Momenten der stationären Verteilung von Interesse. Besonders hervorzuheben ist hierbei die grundlegende und weitreichende Arbeit von Tweedie [33] - „The existence of moments for stationary markov chains“, die für die vorliegende Arbeit insbesondere über die Anwendung in Feigin und Tweedie [32] an Bedeutung gewinnt. Wenn die allgemeinen Ergebnisse zur Existenz von Momenten einer stationären Verteilung aus [33] in den Hintergrund rücken und eher die Untersuchung der Tails im Vordergrund steht, sind nennenswerte Beispiele für die Anwendung von [33] unter anderem Borkovec und Klüppelberg [19], sowie Diebolt und Guégan [11]. Hierbei hebt die letztgenannte Arbeit nicht nur die Wichtigkeit hervor: „The existence of certain exponential moments and knowledge of sharp upper bounds for the tails of the stationary distribution of (1.1) are of primary importance ...“, sondern stellt die enge Beziehung zwischen den Momenten des Störterms und den Momenten der stationären Verteilung mit Proposition 2 - [11] heraus. In diesem Fall (ARCH(1) inbegriffen) überträgt sich die Existenz der polynomialen Momente des Störterms auf die stationäre Verteilung, unter der zusätzlichen Forderung des eingeschränkten Wertebereichs  $(0,1)$  für eine einzige Konstante. Ein solch starker Zusammenhang zu den Momenten des Störterms soll insbesondere auch mittels des folgenden Beispiels verdeutlicht werden.

In der Arbeit von BuK [19], wird die Existenz des zweiten Moments der stationären Verteilung des AR(1)-ARCH(1)-Modells gesichert. Dies geschieht über die gleiche Driftbedingung wie oben  $(\beta^2 + \alpha_1 \mathbb{E}(\varepsilon^2) < 1)$ , jedoch mit dem „small“-set  $A := \left\{ w \in \mathbb{R} \mid w^2 \leq \max \left\{ 1, \frac{\alpha_0 \mathbb{E}(\varepsilon^2)}{(1-2\delta) - (\beta^2 + \alpha_1 \mathbb{E}(\varepsilon^2))} \right\} \right\}$ , wobei  $\delta > 0$  so dass  $(1 - 2\delta) - (\beta^2 + \alpha_1 \mathbb{E}(\varepsilon^2)) > 0$ . Mit der Festlegung von  $h = 1 + w^2$  und unter Zuhilfenahme von Theorem 3 (Tweedie [33] - vgl. Feigin und Tweedie [32]).

Der Unterschied zu den vorangegangenen Überlegungen, zum Überführen der Driftbedingung vom unglätteten zum glätteten Fall und dem obigen Beweis, besteht in der Konstruktion von  $A$ . Hier hat man es mit einer ganz konkreten und von den Parametern definierten Gestalt für  $A$  zu tun, während zuvor eine etwas schwammige Festlegung eines hinreichend großen  $M'$ s genügte. Die Quintessenz des Beweises ist jedoch wiederum die, dass man für alle  $w$ 's, größer oder gleich einer wie auch immer genau definierten Schranke, eine Abschätzung nach oben erhält - vgl. oben bzw. Beweis zu Theorem 1(c) in BuK [19].

In Anlehnung an Example 2 in Feigin und Tweedie [32], lässt sich diese Idee auch auf höhere Momente anwenden, auch wenn die notwendigen Bedingungen dann nicht mehr so überschaubar bzw. verhältnismässig leicht zu gewährleisten sind. Durch diese Vorgehensweise lässt sich aber sehr gut der enge Zusammenhang zwischen den Momenten der stationären Verteilung und den Momenten des Störterms verdeutlichen.

Genauer gesagt geht es nicht um beliebige Momente, sondern jeweils um gerade Momente und veranschaulichen soll dies der Fall  $h = 1 + w^4$ , für  $w \in A^c := \text{s.u.}$



$$\begin{aligned}
& \int_{\mathbb{R}} h(y) P(W_1 \in dy \mid W_0 = w) \\
&= 1 + \mathbb{E}((\beta w + \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon)^4) \\
&= 1 + \mathbb{E}(\beta^4 w^4 + 2\beta w \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon \beta^2 w^2 + (\alpha_0 + \alpha_1 w^2) \varepsilon^2 \beta^2 w^2 + \beta^2 w^2 2\beta w \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon \\
&\quad + 4\beta^2 w^2 (\alpha_0 + \alpha_1 w^2) \varepsilon^2 + (\alpha_0 + \alpha_1 w^2) \varepsilon^2 2\beta w \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon + \beta^2 w^2 (\alpha_0 + \alpha_1 w^2) \varepsilon^2 \\
&\quad + 2\beta w \sqrt{\alpha_0 + \alpha_1 w^2} \varepsilon (\alpha_0 + \alpha_1 w^2) \varepsilon^2 + (\alpha_0 + \alpha_1 w^2)^2 \varepsilon^4) \\
&= 1 + w^4 \mathbb{E}(\beta^4 + 6\beta^2 \alpha_1 \varepsilon^2 + 4\beta(\alpha_1 + \frac{\alpha_0}{w^2})^{\frac{3}{2}} \varepsilon^3 + \alpha_1^2 \varepsilon^4 + \frac{6\beta^2 \alpha_0 \varepsilon^2}{w^2} + \frac{2\alpha_1 \alpha_0 \varepsilon^4}{w^2} + \frac{\alpha_0^2 \varepsilon^4}{w^4}) \\
&\leq 1 + w^4 \mathbb{E}(\beta^4 + 6\beta^2 \alpha_1 \varepsilon^2 + 4\beta(\alpha_1 + \alpha_0)^{\frac{3}{2}} \varepsilon^3 + \alpha_1^2 \varepsilon^4 + \frac{6\beta^2 \alpha_0 \varepsilon^2}{w^2} + \frac{2\alpha_1 \alpha_0 \varepsilon^4}{w^2} + \frac{\alpha_0^2 \varepsilon^4}{w^2}) \\
&\leq 1 + w^4 \left( \beta^4 + 6\beta^2 \alpha_1 \mathbb{E}(\varepsilon^2) + 4\beta(\alpha_1 + \alpha_0)^{\frac{3}{2}} \mathbb{E}(\varepsilon^3) + \alpha_1^2 \mathbb{E}(\varepsilon^4) + \frac{6\beta^2 \alpha_0 \mathbb{E}(\varepsilon^2)}{w^2} + \frac{2\alpha_1 \alpha_0 \mathbb{E}(\varepsilon^4)}{w^2} + \frac{\alpha_0^2 \mathbb{E}(\varepsilon^4)}{w^2} \right) \\
&\leq 1 + w^4(1 - 2\delta) = 1 - w^4 \delta + w^4(1 - \delta) \leq 1 - \delta + w^4(1 - \delta) \\
&= h(w)(1 - \delta)
\end{aligned}$$

Hierbei ergibt sich die letzte Zeile aus der Forderung

$$\begin{aligned}
& (1 - 2\delta) > \beta^4 + 6\beta^2 \alpha_1 \mathbb{E}(\varepsilon^2) + 4\beta(\alpha_0 + \alpha_1)^{\frac{3}{2}} \mathbb{E}(\varepsilon^3) + \alpha_1^2 \mathbb{E}(\varepsilon^4) + \frac{6\beta^2 \alpha_0 \mathbb{E}(\varepsilon^2) + (2\alpha_1 \alpha_0 + \alpha_0) \mathbb{E}(\varepsilon^4)}{w^2} \\
&\Leftrightarrow w^2(1 - 2\delta) > w^2(\beta^4 + 6\beta^2 \alpha_1 \mathbb{E}(\varepsilon^2) + 4\beta(\alpha_1 + \alpha_0)^{\frac{3}{2}} \mathbb{E}(\varepsilon^3) + \alpha_1^2 \mathbb{E}(\varepsilon^4)) + 6\beta \alpha_0 \mathbb{E}(\varepsilon^2) + (2\alpha_1 \alpha_0 + \alpha_0^2) \mathbb{E}(\varepsilon^4) \\
&\Leftrightarrow w^2((1 - 2\delta) - (\beta^4 + 6\beta^2 \alpha_1 \mathbb{E}(\varepsilon^2) + 4\beta(\alpha_1 + \alpha_0)^{\frac{3}{2}} \mathbb{E}(\varepsilon^3) + \alpha_1^2 \mathbb{E}(\varepsilon^4))) > 6\beta \alpha_0 \mathbb{E}(\varepsilon^2) + (2\alpha_1 \alpha_0 + \alpha_0^2) \mathbb{E}(\varepsilon^4) \\
&\Leftrightarrow w^2 > \frac{6\beta \alpha_0 \mathbb{E}(\varepsilon^2) + (2\alpha_1 \alpha_0 + \alpha_0^2) \mathbb{E}(\varepsilon^4)}{(1 - 2\delta) - (\beta^4 + 6\beta^2 \alpha_1 \mathbb{E}(\varepsilon^2) + 4\beta(\alpha_1 + \alpha_0)^{\frac{3}{2}} \mathbb{E}(\varepsilon^3) + \alpha_1^2 \mathbb{E}(\varepsilon^4))}
\end{aligned}$$

und damit die Gestalt von

$$A := \left\{ w \in \mathbb{R} \mid w^2 \leq \max \left\{ 1, \frac{6\beta \alpha_0 \mathbb{E}(\varepsilon^2) + (2\alpha_1 \alpha_0 + \alpha_0^2) \mathbb{E}(\varepsilon^4)}{(1 - 2\delta) - (\beta^4 + 6\beta^2 \alpha_1 \mathbb{E}(\varepsilon^2) + 4\beta(\alpha_1 + \alpha_0)^{\frac{3}{2}} \mathbb{E}(\varepsilon^3) + \alpha_1^2 \mathbb{E}(\varepsilon^4))} \right\} \right\}.$$

Damit diese Konstruktion bzw. Beweisidee funktioniert, muss gefordert werden

$(1 - 2\delta) - (\beta^4 + 6\beta^2 \alpha_1 \mathbb{E}(\varepsilon^2) + 4\beta(\alpha_1 + \alpha_0)^{\frac{3}{2}} \mathbb{E}(\varepsilon^3) + \alpha_1^2 \mathbb{E}(\varepsilon^4)) > 0$  und als Hauptbedingung ergibt sich  $\mathbb{E}|\beta + \sqrt{\alpha_0 + \alpha_1} \varepsilon|^4 < 1$ .

Verfolgt man diesen Ansatz konsequent weiter, so ergibt sich als hinreichende Bedingung für die Existenz des k-ten Moments (für k gerade,  $k = 2m$ ,  $m \in \mathbb{N}$ )

$$\sum_{n=0}^k \binom{k}{n} \beta^{k-n} (\alpha_0 + \alpha_1)^{\frac{n}{2}} \mathbb{E}(\varepsilon^n) < 1.$$

Während für das zweite Moment noch einiges zusammenfällt und einen überschaubaren Ausdruck entstehen lässt, kann man an der länglichen Bedingung fürs vierte Moment zumindest ablesen, dass die Existenz der Momente des Störterms einen entscheidenden Einfluss hat.

Auch wenn in der vorliegenden Arbeit keine weiteren Untersuchungen der Tails unternommen werden, macht allein die Verwendung der Ergebnisse von Tweedie [33] bzw. von Feigin und Tweedie [32] eine Erwähnung sinnvoll, um deren Tragweite zu erkennen. Eine Anknüpfung an die bisherigen Ergebnisse ist also, trotz unterschiedlicher Zielsetzung, über das verwendete Instrumentarium möglich.

### 3 Empirische Studie

Dieses Kapitel soll mittels einer empirischen Studie des Kurses der Allianz-Aktie die bisherigen Ergebnisse erweitern bzw. untermauern. Obwohl die historischen Kursdaten bis zum heutigen Tage vorliegen, soll der Betrachtungszeitraum zunächst eingeschränkt werden auf das Zeitfenster 02.01.1990 - 30.12.2004, um so die Ergebnisse mit den Werten in Hoyer [15] vergleichen zu können. Durch diese Vorgehensweise kann man sich von der korrekten Arbeitsweise der bereits vorhandenen Algorithmen, aber auch eigener Routinen überzeugen.

Die Quelle für das entsprechende Datenmaterial lautet <http://www.ariva.de/chart/?secu=292> und unterscheidet sich somit zunächst von der in Hoyer [15] verwendeten Quelle Bloomberg. Nach einer ersten Aufarbeitung mit OpenOffice Calc zeigt sich zunächst, dass die Stichprobenanzahl mit 3911 um 137 Datenpunkte höher ist als das zugrunde gelegte Datenmaterial bei Hoyer [15]. Der Unterschied erklärt sich durch eine bei Hoyer vorgenommene Feiertagsbereinigung, während bei dem hier genutzten Datenmaterial, die durch Feiertage entstandenen Lücken durch Interpolation gefüllt wurden. Aufgrund der verhältnismässig geringen Anzahl von zusätzlichen Datenpunkten, soll auf eine Anpassung verzichtet werden. Des Weiteren ist mit der abweichenden Datenquelle und Verwendung des kostenlosen Softwarepaketes R, anstelle von Matlab, ohnehin nicht mit einer 100%igen Übereinstimmung der Ergebnisse zu rechnen.

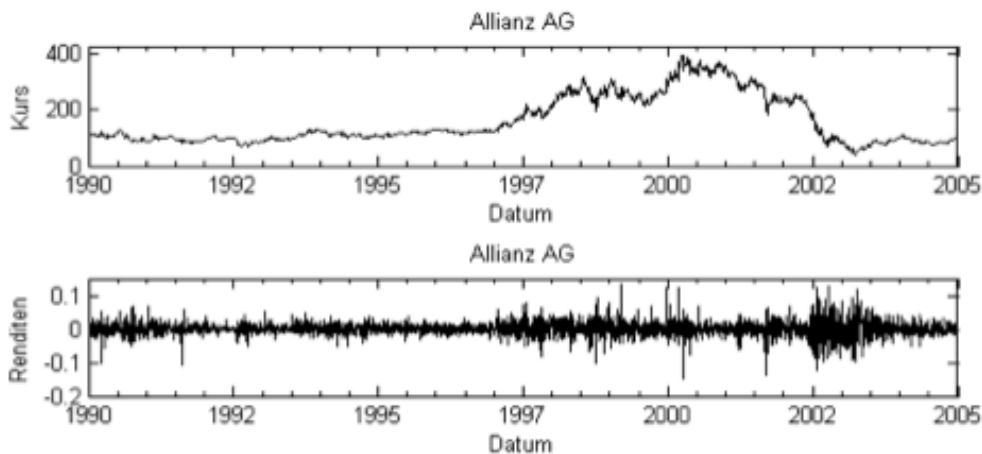


Abbildung 3.1: Allianz Kurs und Renditen (Bloomberg vgl. [15])



Abbildung 3.2: Allianz Kurs (www.ariva.de)

Ein grafischer Vergleich legt jedoch nahe, dass die Ergebnisse erkennbar ähnlich bzw. nah beieinander liegend sind - Abbildung 3.2 ist Grundlage für unsere Untersuchung, während Abbildung 3.1 aus [15] stammt. Die Renditen oder Log>Returns werden im nächsten Abschnitt definiert

und erläutert.

Für die Konstruktion eines AR(1)-ARCH(1)-Modells bedient man sich einiger wichtiger Bausteine, wie etwa dem grundlegenden White-Noise-Prozess. Bevor man mit der Parameterschätzung beginnt, wäre es empfehlenswert sich die einzelnen Prozesse verbunden mit dem entsprechenden Programmcode anzuschauen. Dadurch sind deren Verwendung bzw. Umsetzung in der Statistiksoftware R und die kommenden Darstellungen wesentlich schneller zu verstehen (siehe Anhang - Abschnitt 4.5).

Zielsetzung soll es sein, mit Hilfe von R einen Algorithmus zu schreiben, der nicht nur den AR(1)-ARCH(1)-Fall abdeckt, sondern auch für den geglätteten Ansatz eine Parameterschätzung ermöglicht.

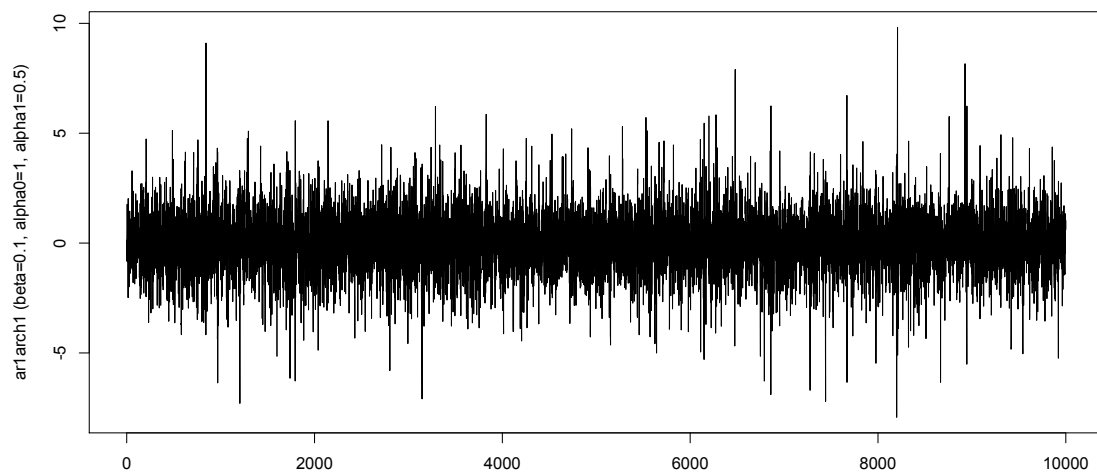
### 3.1 Simulation AR(1)-ARCH(1)

Als Kombination der zugrundeliegenden Prozesse und somit als Sammelwerk ihrer besonderen Eigenschaften, erhält man den AR(1)-Prozess mit ARCH(1)-Fehlerterm oder kurz den AR(1)-ARCH(1)-Prozess.

Simulation AR(1)-ARCH(1)

$$W_{t+1} = \beta W_t + \sqrt{\alpha_0 + \alpha_1 W_t^2} \varepsilon_{t+1}$$

```
> w<-c(1:10000)
> w<-ts(w)
> beta<-0.5
> alpha0<-0.6
> alpha1<-0.7
> for(i in 1:9999){
+ w[i+1]<-beta*w[i] + sqrt(alpha0 + alpha1*(w[i]^2))*eps[i+1]}
> ar1arch1<-w
```



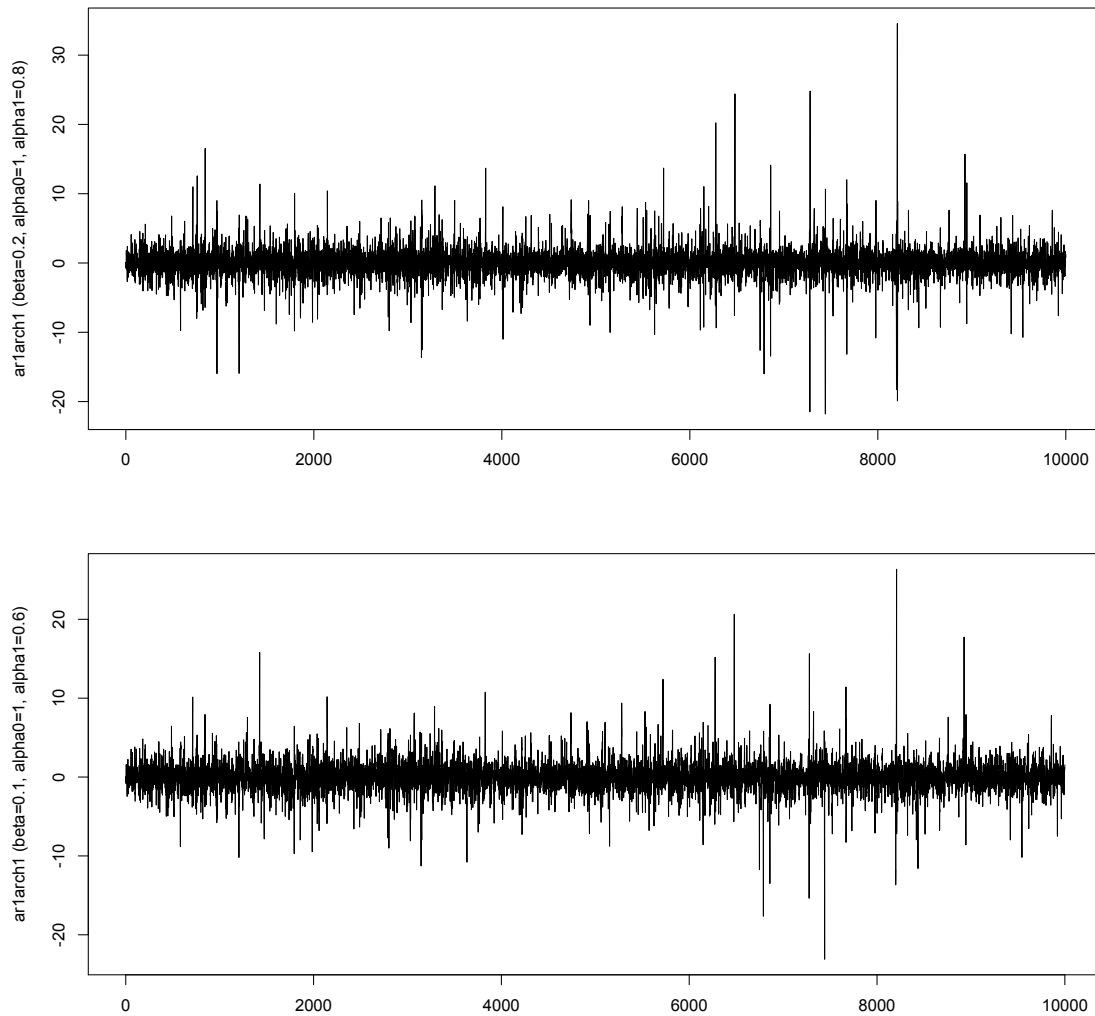


Abbildung 3.3: Drei Realisationen eines AR(1)-ARCH(1)-Prozesses bei identischem Störterm.

Für dieses Beispiel und die im Anhang befindlichen Beispiele (Abschnitt 4.5), könnte man die Parameter als Argumente in den Funktionsaufruf integrieren oder auch die Länge des zu betrachtenden Zeitraums variabel machen. Zum Zwecke einer Simulation wäre es aber auch möglich, auf eine eigene Implementierung zu verzichten und stattdessen auf bereits in R verfügbare Algorithmen zurückzugreifen (vgl. R-Pakete `fseries` oder `fGarch`).

Die Vorteile einer eigenen Implementierung liegen jedoch auf der Hand, da man bei dieser Vorgehensweise alle Parameter kontrollieren, den Prozess in jedem Entstehungsschritt verfolgen und von daher viel besser im Detail verstehen kann. So liegt im AR(1)<sup>6</sup>, ARCH(1)<sup>7</sup> und AR(1)-ARCH(1)-Beispiel nicht nur eine i.i.d. Folge mit gleicher Verteilung als Störterm zu Grunde, sondern bei identischem Startwert für den Zufallsgenerator `set.seed(.)` (siehe `White.Noise`) auch identische Störgrößen, das heißt:  $\epsilon_{\text{ps}}(i)$ ,  $i \in \{1, \dots, 10000\}$  stimmt in jeder dieser Realisationen überein. Alle Unterschiede können also auf die verschiedenen Konstruktionen bzw. Parameter zurückgeführt werden und sind nicht von unterschiedlichen Störgrößen abhängig. Dies ist im Sinne einer Reproduzierbarkeit und tieferem Verständnis von immenssem Vorteil.

<sup>6</sup>siehe Anhang - Abschnitt 4.5

<sup>7</sup>siehe Anhang - Abschnitt 4.5

Computergenerierte Zufallszahlen sind nämlich keineswegs „zufällig“, sondern werden mittels verschiedener Algorithmen generiert. Die Erzeugung von Zufallszahlen ist der zentrale Punkt in Kolonko [20] und bietet also die Grundlage für obigen Vergleich. Darüber hinaus helfen einem vorgefertigte Pakete bei der Umsetzung eines noch nicht in R enthaltenen Beispiels, wie dem geglätteten AR(1)-ARCH(1)-Prozess, nur bedingt weiter.

### 3.2 Simulation Smooth-AR(1)-ARCH(1)

Der geglättete AR(1)-ARCH(1)-Prozess<sup>8</sup> entsteht, entsprechend dem oben bereits gezeigten Schema, indem man zunächst den Wert des AR(1)-ARCH(1)-Prozesses bildet und anschließend bei Bedarf glättet. Dieser gegebenenfalls geglättete Wert geht in die nächste Iteration des AR(1)-ARCH(1)-Prozesses ein und anschließend wird bei Bedarf wieder geglättet. Für die Umsetzung bedeutet dies, dass wir die Simulation nur um die Glättungskomponente zu erweitern brauchen. Da die Berechnung rekursiv mit einer For-Schleife geschieht, muss die Glättungsfunktion in diese Schleife eingebunden werden. Eine solche Vorgehensweise wäre in einem vorgefertigten Simulationsalgorithmus nicht möglich. Wir beschränken uns zunächst wieder auf den einfachen Fall einer Glättung mit Geradenstücken - vgl. Definition  $g^{**}(\cdot)$  s.o. - und müssen demnach noch die Glättungsparameter  $\alpha$ ,  $\gamma$  sowie die Schranken  $\underline{w}_0$ ,  $\overline{w}_0$  mit einbeziehen. Die zwei Parameter  $\alpha$ ,  $\gamma$  dienen, wie bereits erwähnt, als Stellschrauben für die Stärke der Glättungsfunktion und die zwei Schranken  $\underline{w}_0$ ,  $\overline{w}_0$  legen das Kompaktum fest, in dem nicht geglättet wird.

Simulation Smooth-AR(1)-ARCH(1)

$$W_{t+1} = g^{**}(\beta W_t + \sqrt{\alpha_0 + \alpha_1 W_t^2} \varepsilon_{t+1})$$

```
> w<-c(1:10000)
> w<-ts(w)
> beta<-0.1
> alpha0<-1
> alpha1<-0.5
# Glättungsparameter
> alpha<-0.2
> gamma<-0.3
# Schranken fürs Kompaktum
> w0unten<- -5
> w0oben<- 5
> for(i in 1:9999){
+ w[i+1]<-beta*w[i] + sqrt(alpha0 + alpha1*(w[i]^2))*eps[i+1]
# if-Abfrage und Anweisung für Wert >  $\overline{w}_0$ 
+ if(w[i+1]>w0oben){
+ w[i+1]<-w0oben + gamma*(w[i+1] - w0oben)
}
# if-Abfrage und Anweisung für Wert <  $\underline{w}_0$ 
+ if(w[i+1]<w0unten){
```

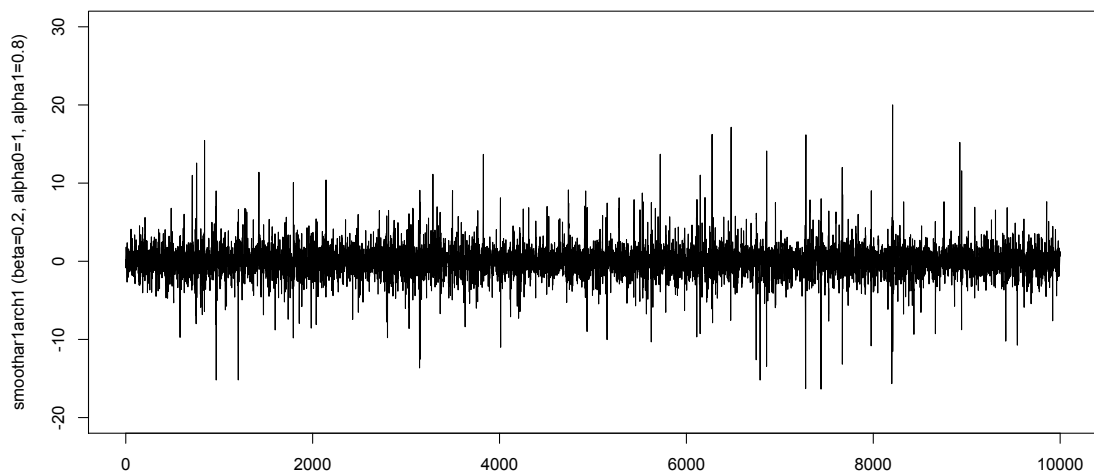
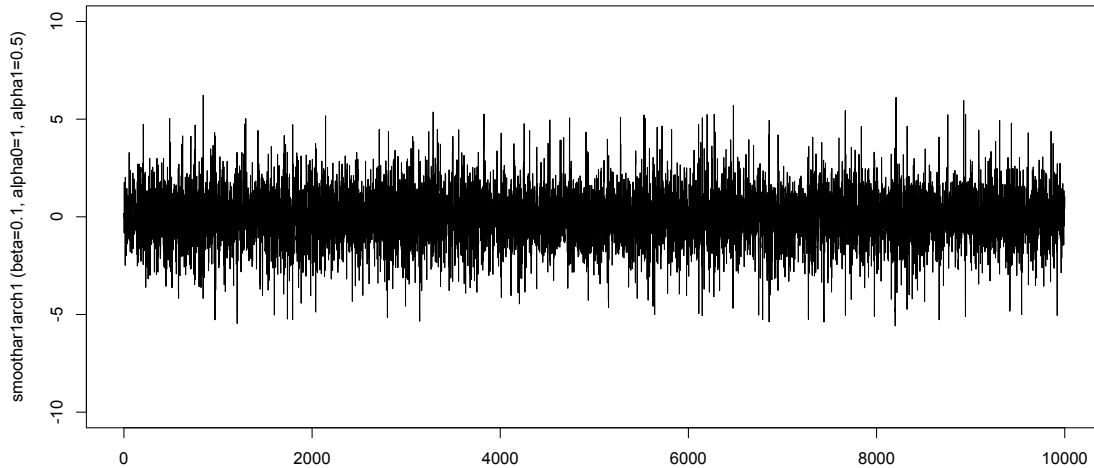
---

<sup>8</sup>im folgenden auch Smooth-AR(1)-ARCH(1) genannt

```

+ w[i+1]<-w0unten + alpha*(w[i+1] - w0unten)
}
# Sofern keine if-Abfrage positiv ist, bleibt der Wert erhalten und die Schleife
startet neu
> w[i+1]<-w[i+1]
}
> smoothar1arch1<-w

```



Der Graph zeigt, wie sich der Glättungseffekt auswirkt. Die in den drei AR(1)-ARCH(1)-Simulationen typischen Ausreißer nach oben und nach unten, werden nicht radikal abgeschnitten, aber abgeschwächt. Die Werte sind immer noch größer als die vorgegebene Schranke, überragen die restlichen Werte aber nicht mehr so eindeutig. Anders als beim Abschneiden, bleibt das Verhältnis der Ausreißer auch untereinander erhalten. Der größte oder kleinste Ausreißer bleibt also auch nach dem Glätten als solcher erkennbar, wird aber in seiner Auswirkung auf die nachfolgenden Werte abgeschwächt. In Hinblick auf das grundlegende Verhalten eines solchen Prozesses, würden die sehr großen und sehr kleinen (negativen) Werte das Bild ansonsten stark verzerren.

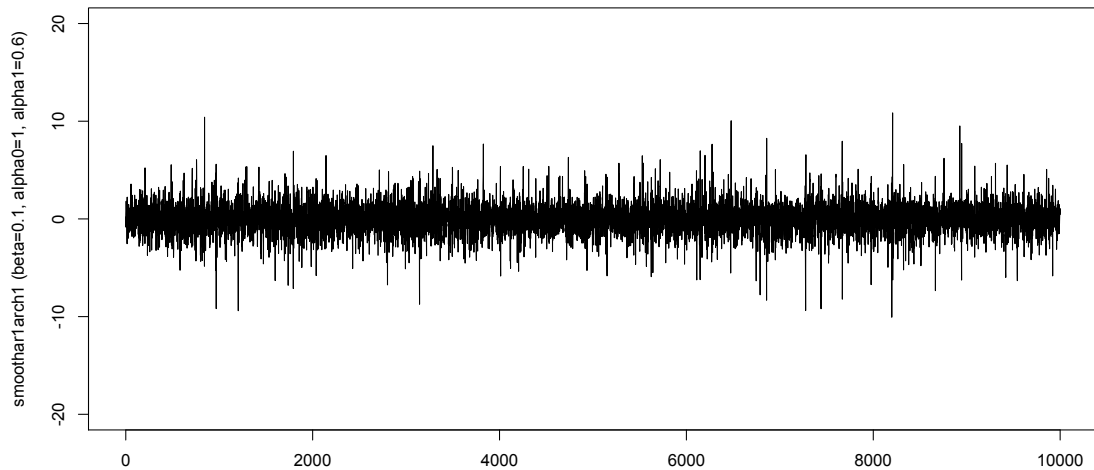


Abbildung 3.4: Smooth-AR(1)-ARCH(1)-Realisationen mit  $\alpha = 0.2, \gamma = 0.3, \underline{w}_0 = -5, \overline{w}_0 = 5, \underline{w}_0 = -15, \overline{w}_0 = 15$  bzw.  $\underline{w}_0 = -10, \overline{w}_0 = 10$ . Störterm identisch wie in Abbildung 3.3.

### 3.3 Simulation Poly-Smooth-AR(1)-ARCH(1)

Als letztes Beispiel wird der Polynomiale Glättungsansatz behandelt. Obwohl die Änderung von den Vorfaktoren  $\alpha, \gamma$  zu den entsprechenden Exponenten  $\alpha, \gamma$  sehr harmlos wirkt, sind die oben bereits thematisierten Effekte äußerst weitreichend. In Hinblick auf den Simulationsalgorithmus kann man sich aber der simplen Änderung erfreuen.

Simulation Poly-Smooth-AR(1)-ARCH(1)

$$W_{t+1} = g^{4*}(\beta W_t + \sqrt{\alpha_0 + \alpha_1 W_t^2} \varepsilon_{t+1})$$

```
> w<-c(1:10000)
> w<-ts(w)
> beta<-0.1
> alpha0<-1
> alpha1<-0.5
# Glättungsparameter
> alpha<-0.2
> gamma<-0.3
# Schranken fürs Kompaktum
> w0unten<- -5
> w0oben<- 5
> for(i in 1:9999){
+ w[i+1]<-beta*w[i] + sqrt(alpha0 + alpha1*(w[i]^2))*eps[i+1]
# if-Abfrage und Anweisung für Wert >  $\overline{w}_0$ 
+ if(w[i+1]>w0oben){
+ w[i+1]<-(w0oben-1) + (w[i+1] - w0oben + 1)^gamma
}
```

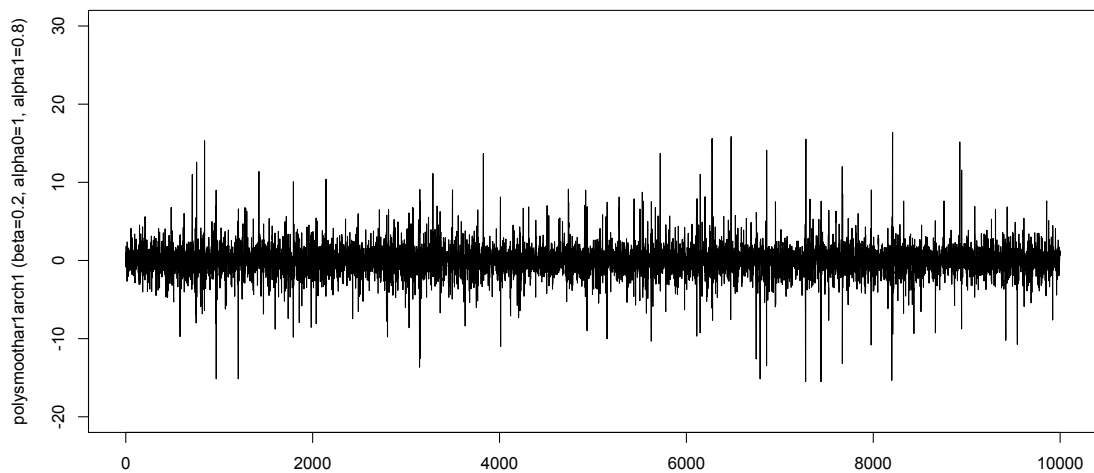
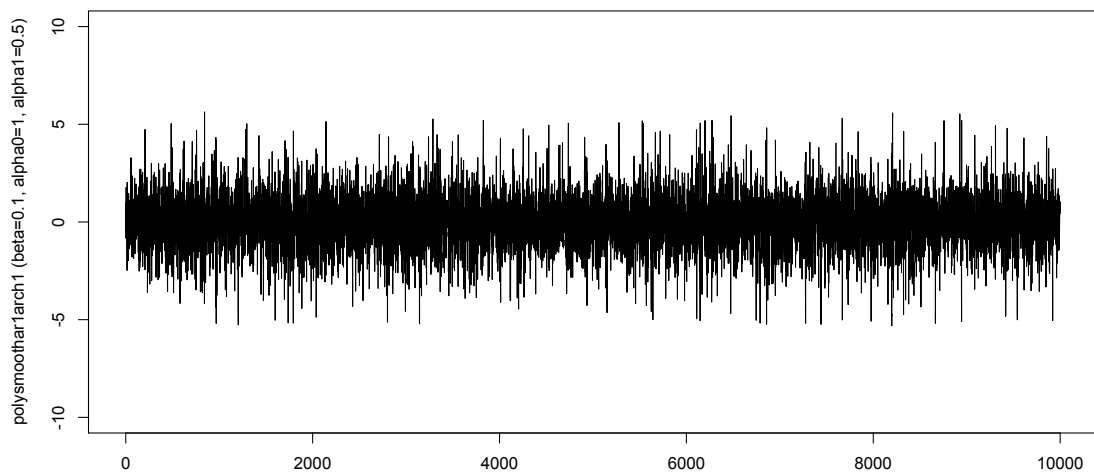


```

}
# if-Abfrage und Anweisung für Wert <  $w_0$ 
+ if(w[i+1]<w0unten){
+ w[i+1]<-(w0unten+1) + (w0unten + 1 - w[i+1])^alpha
+ }
# Sofern keine if-Abfrage positiv ist, bleibt der Wert erhalten und die Schleife
startet neu
> w[i+1]<-w[i+1]
}
> polysmoothar1arch1<-w

```

Der Unterschied zum Geradenbeispiel zeigt sich im Algorithmus nur in den beiden `if`-Abfragen und der tatsächliche Effekt, lässt sich wiederum an der Grafik veranschaulichen.



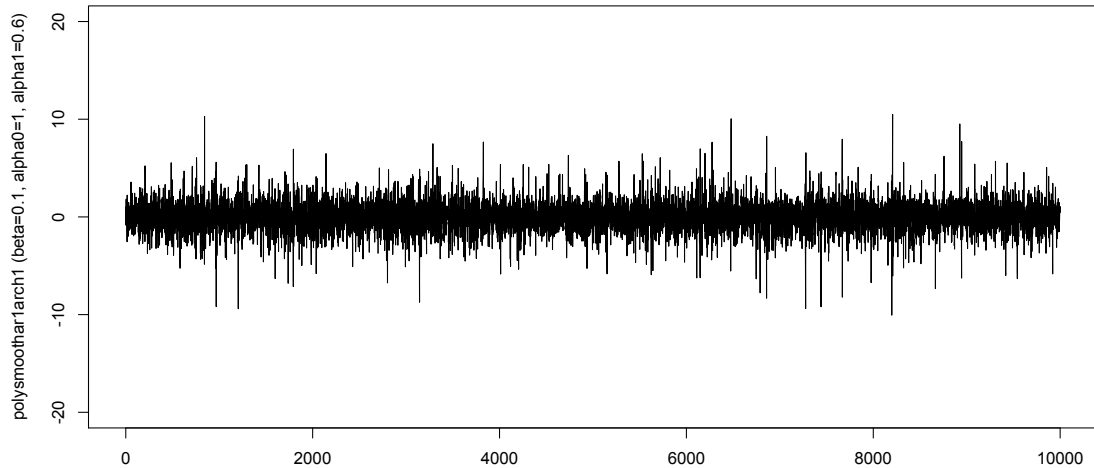


Abbildung 3.5: Poly-Smooth-AR(1)-ARCH(1)-Realisationen mit  $\alpha = 0.2, \gamma = 0.3, \underline{w}_0 = -5, \overline{w}_0 = 5, \underline{w}_0 = -15, \overline{w}_0 = 15$  bzw.  $\underline{w}_0 = -10, \overline{w}_0 = 10$ . Störterm identisch wie in Abbildung 3.3.

Wie durch die Konstruktion und die theoretischen Überlegungen nahe gelegt, zeigt sich der Unterschied umso stärker, je weiter man sich von den Startpunkten für die Glättung  $\underline{w}_0, \overline{w}_0$  entfernt. Am ehesten sieht man den Effekt bei der jeweils mittleren Abbildung ( $\alpha_1 = 0.8$ ), wenn man sich auf den großen positiven Ausschlag bei ca.  $t=8200$  konzentriert. Man erkennt an dieser Stelle wunderbar, wie der „größte“ Ausreißer durch die Glättung eingefangen wird, aber weiterhin als „größter“ Wert erhalten bleibt. Des weiteren kann man an diesem Beispiel auch den stärkeren Glättungseffekt des polynomialen Ansatzes sehen, da der Ausschlag in diesem Fall deutlich kleiner ausfällt, als bei der Geradenglättung mit den gleichen Parametern (vgl. obige Herleitung und Diskussion der Glättungsansätze).

Hat man sich einmal mit den Besonderheiten der unterschiedlichen Glättungsansätze vertraut gemacht, muss man sich früher oder später auch Gedanken darüber machen, ob sich die ganze Mühe denn gelohnt hat. Es gilt also die Frage zu beantworten, ob man aus dieser Herangehensweise auch einen Nutzen ziehen kann. Hintergrund der gesamten Glättungsüberlegungen ist eine nach Möglichkeit verbesserte Anpassung an gegebenes Datenmaterial. In diesem Sinne besteht der nächste Schritt darin, sich mit der Parameterschätzung auseinanderzusetzen.

### 3.4 Parameterschätzung

Schaut man sich die Log>Returns des Allianz-Beispiels in Abbildung 3.6 genau an, so stellt man fest, dass nur eine Hand voll Werte über 0.1 oder unter -0.1 liegen. Rein intuitiv könnte man also annehmen, dass es mit einem Modell, welches solch starke Ausreißer nur bedingt zulässt, zu einer besseren Anpassung an diese Daten kommt. Dies ist an dieser Stelle natürlich nur eine Vermutung, doch insbesondere die Modellauswahl ist zumeist nicht auf Grundlage von harten Fakten und Kennziffern zu gestalten, sondern beruht sehr wohl auf Erfahrung und Intuition. Eine für Aktienkurse typische und klar erkennbare Clusterbildung<sup>9</sup> würde mit der Erfahrung in solchen Fällen auf die Klasse der ARCH und GARCH-Modelle hindeuten, jedoch ist meist erst am Endergebnis ablesbar, ob die getroffene Wahl oder die Parameter als zutreffend zu bezeichnen sind.

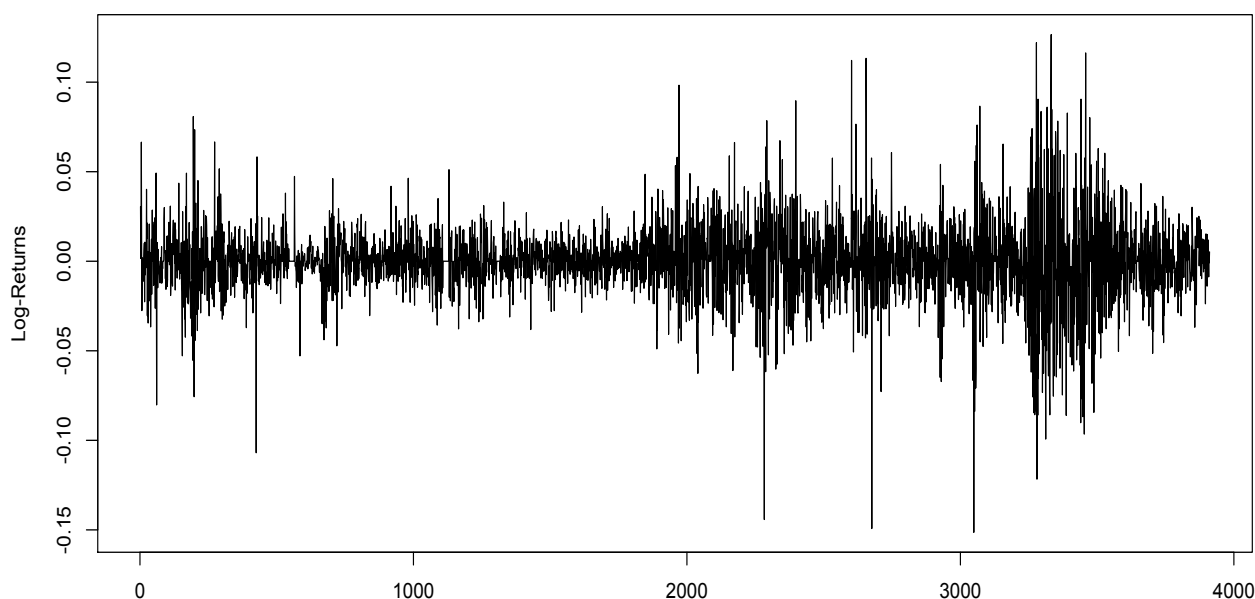


Abbildung 3.6: Log>Returns zum Allianz Kurs (Abb. 3.2)

Das R-Paket `fGarch` enthält unter anderem eine Prozedur `garchFit`, mit der man für sehr allgemeine Modelle eine Parameterschätzung vornehmen kann. Wie bereits erwähnt, ist nicht damit zu rechnen, dass wir zu exakt den gleichen Werten kommen, wie sie bei Hoyer [15] errechnet wurden. Andererseits sind wir dank obiger Simulationen in der Lage, etwas über die Präzision der Schätzungen von `garchFit` auszusagen.

Während AR(1)-ARCH(1) durch `garchFit` abgedeckt wird, muss für die geglätteten AR(1)-ARCH(1)-Fälle jeweils eine eigene Prozedur implementiert werden und auch deren Wirksamkeit wird sich erst im Vergleich zu Ergebnissen von `garchFit` beurteilen lassen. Obige Simulationen sind demnach keine Zeitverschwendung oder vernachlässigbar, sondern entscheidend für weitere Beurteilungen.

Die Daten des Allianz-Kurses (hier mit  $z$  bezeichnet) werden jedoch nicht in dieser Ursprungsform zugrunde gelegt, sondern zunächst logarithmiert ( $y_t = \log z_t$ ) und dann mit einem Diffe-

---

<sup>9</sup>vgl. z.B. Straumann [28] - Stichwort: „Stylized facts of financial data“

renzenfilter

$$x_t = y_{t+1} - y_t =: \Delta y_{t+1}$$

behandelt - siehe etwa [10] S.242 oder [27] S.39. Dies führt zu den sogenannten Log>Returns<sup>10</sup> und erst mit diesen Daten startet man die Parameterschätzung, d.h die Log>Returns werden als AR(1)-ARCH(1) modelliert. In Fan [37] - S.168 wird die Motivation zu derartigen Umformungen genauer umrissen und auch mögliche Abwandlungen angesprochen. Weitere Möglichkeiten zur Transformation der Daten, werden auch in Schlittgen [29] S.35ff beschrieben. So sind die (Log)-Returns von  $\{Z_t\}$  typischerweise definiert durch

$$X_t = \log Z_t - \log Z_{t-1} \quad \text{oder} \quad Y_t = \frac{Z_t - Z_{t-1}}{Z_{t-1}}$$

und unterscheiden sich nur unwesentlich, da

$$\log Z_t - \log Z_{t-1} = \log \left( 1 + \frac{Z_t - Z_{t-1}}{Z_{t-1}} \right) \approx (Z_t - Z_{t-1})/Z_{t-1}.$$

Mal abgesehen von Lehrbüchern zur Zeitreihenanalyse [37], [10], [31], [29], [27] und Vorlesungsskripten aus dem Bereich [39], werden solche Transformation der Daten meist gar nicht genauer erläutert, sondern als Notwendigkeit hingenommen. Beispiele für diese Vorgehensweise sind BuK [19], Truquet [30] oder Francq und Zakoian [38], die entweder gänzlich ohne Hinweis auskommen oder „return series“ nur als übliches Anwendungsgebiet nennen. Je nach Ausgangssituation kann sich die genaue Methode als auch deren Verwendungshäufigkeit stark unterscheiden. Die Idealwelt, wie man sie in Lehrbüchern vorfindet, hat man bei einem konkreten Problem natürlich nicht vor sich, aber man erkennt unschwer, dass der Aktienkurs nicht viel mit den Schaubildern der oben durchgeführten Simulationen gemein hat. Der Allianz-Kurs hat über längere Passagen einen Trend nach oben oder nach unten. Solch ein Trend widerspricht jedoch der Stationaritätsannahme (für den Erwartungswert), die man im Rahmen einer Zeitreihenanalyse macht. Die Differenzenbildung kann diesen Trend eliminieren. Bei sich über die Zeit verändernder Größenordnung der Ausschläge, kann hingegen das Logarithmieren helfen (mögliche Instationarität der Varianz), sofern natürlich nur positive Werte vorhanden sind. Für Prognosen sind solche Stationaritätseigenschaften jedoch entscheidend und deshalb obige Transformationen meist einfach selbstverständlich. Diese Vorgehensweise dient also dem Zwecke der „Stationarisierung“ und ist in R ohne größeren Aufwand durchführbar:

```
> x<-diff(log(Allianz))
```

Die Anpassung der Allianz Log>Returns an ein AR(1)-ARCH(1)-Modell, führt mit den Daten von Bloomberg und der Matlab basierten Prozedur `garchFit` zu den folgenden Ergebnissen:

AR(1)-ARCH(1)

$\alpha_0 = 0.00032303$ ,  $\alpha_1 = 0.3357$ ,  $\beta = 0.0966$

Im Vergleich dazu, die Anpassung der Allianz Log>Returns aus den Ariva-Daten, mit der R basierten `garchFit` Prozedur:

---

<sup>10</sup>auch Renditen genannt - vgl. Graphiken

AR(1)-ARCH(1)

$\alpha_0 = 0.0002792524$ ,  $\alpha_1 = 0.4030117572$ ,  $\beta = 0.1379416664$

Besonders vor dem Hintergrund der unterschiedlichen Stichprobenlängen, Datenquellen und der verschiedenen Schätzprozeduren, erscheinen die Abweichungen durchaus akzeptabel. Selbst bei absolut identischen Datensätzen kann man, je nach Algorithmus, immer noch zu unterschiedlichen Ergebnissen gelangen (vgl. [22] für GARCH(1,1)). Alle errechneten Werte sind demnach immer mit einer gewissen Vorsicht zu genießen, insbesondere wenn man gar keine Information zur konkreten Berechnungsvorschrift erhält. Die Probleme mit der „Filter“-Repräsentation sind ein Paradebeispiel in diesem Zusammenhang - siehe Abschnitt 4.6.

Wir wollen fürs weitere Vorgehen die R basierte Prozedur `garchFit` als Vergleichselement hernehmen und untersuchen deren Ergebnisse zunächst an einigen AR(1)-ARCH(1)-Simulationen.

	Simulation	garchFit	Simulation	garchFit	Simulation	garchFit
$\beta$	0.1	0.0935063162	0.2	0.1795206104	0.5	0.19568190
$\alpha_0$	0.2	0.2130327930	0.6	0.6441333151	0.6	0.67258943
$\alpha_1$	0.3	0.2706823812	0.3	0.2780984584	0.7	0.92173912

Insbesondere die ersten beiden Beispiele verdeutlichen, wie gut man sich den tatsächlichen Parametern nähern kann und geben eine Orientierung für selbstgeschriebene Algorithmen. Beispiel 3 jedoch zeigt auch, dass man sich nicht in jedem Fall auf ein hervorgebrachtes Ergebnis verlassen darf. Vielmehr gilt es, die Entstehung zu hinterfragen und gegebenenfalls mit verfeinerten Methoden zu arbeiten. Vor diesem Hintergrund ist auch der Vergleich mit den Ergebnissen von Hoyer [15] lohnenswert, da die Ergebnisse vertrauenswürdiger erscheinen, wenn Sie trotz unterschiedlicher Plattformen (Maple vs. R) nahezu identisch sind.

Die Parameterschätzung erfolgt über die Quasi-Maximum-Likelihood-Methode - vgl. dazu insbesondere [38], [37] S.156ff, [30]. Einen Einblick in die prinzipielle Maximum-Likelihood-Methode liefern sehr anschaulich Georgii [9], Brockwell [6] oder auch Schlittgen [27].

Ausgangspunkt der Parameterschätzung ist dabei folgender Zusammenhang:

Es sei  $f_{1,\theta}$  die Dichte einer Zufallsvariablen  $X_1$  und  $f_{t,\theta}(\cdot|x_{t-1}, \dots, x_1)$  die bedingte Dichte von  $X_t$  unter der Bedingung  $X_{t-1} = x_{t-1}, \dots, X_1 = x_1$ . Die Dichte  $f_\theta$  von  $(X_1, \dots, X_n)$  hat dann die Darstellung

$$f_\theta(x_1, \dots, x_n) = f_{1,\theta}(x_1) \cdot \prod_{t=2}^n f_{t,\theta}(x_t|x_{t-1}, \dots, x_1)$$

und wird als Likelihood-Funktion bezeichnet. Eine detaillierte Begründung zur Verwendung der Dichtefunktion findet man in Schlittgen [27] S.288-289. Gemäß dem Maximum-Likelihood-Prinzip sucht man jene Parameterkombination, die die vorliegende Stichprobe  $x = (x_1, \dots, x_n)$  als Realisation des Prozesses  $\{X_n\}$  mit größter Wahrscheinlichkeit hervorgebracht hat.

$$\theta_{Opt} = \operatorname{argmax}_{\theta \in \Theta} f_\theta(x)$$

Typischerweise stellt sich jedoch das Problem einer endlichen Stichprobe und eines Vergangenheitsbezuges des betrachteten Prozesses (siehe z.B. AR(p)-Prozess), so dass man auf die Quasi-Maximum-Likelihood-Funktion zurückgreift (auch Pseudo-Maximum-Likelihood-Funktion oder Conditional-Maximum-Likelihood-Funktion genannt - siehe z.B. Straumann [28]):

$$L_x(\theta) = \prod_{t=p+1}^n f_{t,\theta}(x_t|x_{t-1}, \dots, x_1)$$

Aus Gründen der besseren numerischen Umsetzung, wird aus dem Maximierungsproblem durch logarithmieren und einen Vorzeichenwechsel ein Minimierungsproblem. Die Umsetzung dieses Prinzips findet sich in den entsprechenden Programmcodes zur Parameterschätzung. Einige Variationsmöglichkeiten dieser Vorgehensweise bietet Truquet [30] und in Fan [37] S.157ff werden neben dem Quasi-Maximum-Likelihood-Prinzip auch exotischere Schätzer besprochen.

Um den geglätteten Fall ebenfalls behandeln zu können, muss man sich von den vorprogrammierten Prozeduren lösen und einen tieferen Einblick in die genaue Arbeitsweise solcher Algorithmen gewinnen. In diesem Sinne dient `garchFit` weiterhin als Referenz, aber die Parameterschätzung soll mit Hilfe eines kompakteren Algorithmus geschehen. Der Programmcode der Prozedur `garchFit`<sup>11</sup> umfasst im DIN A4-Format 16 Seiten. Dies liegt in erster Linie natürlich daran, dass diese Prozedur eine weite Palette an ARMA(m,n)+GARCH/APARCH(p,q)-Modellen bedient. Diese Variationsmöglichkeit führt jedoch dazu, dass im Programmcode immer wieder Passagen auftauchen, die für ein konkretes Problem keine Bedeutung haben und das direkte „lesen“ und verstehen des Codes erschweren. Darüber hinaus ist die Bedeutung bzw. das Zusammenspiel einiger Abschnitte nur schwer nachvollziehbar bzw. auf den ersten Blick oftmals gar nicht zu erkennen.

Spezialisiert man sich jedoch in Hinblick auf das zu betrachtende Modell, so sind die zentralen Schritte einer Parameterschätzung durchaus kompakt und nachvollziehbar darstellbar, wie das `garch11Fit`-Beispiel im Anhang (Abschnitt 4.6) zeigt, welches nur die Parameter für den GARCH(1,1)-Fall schätzt. Der Programmcode wurde hierbei entnommen aus [22] und füllt im DIN A4-Format nur noch eine Seite. Das Grundgerüst der sechs Hauptschritte wollen wir für unsere weitere Betrachtung übernehmen, jedoch hat das Programm auch seine Schwächen, die man auf den ersten Blick leider kaum erkennen kann. So führt die „Filter“-Repräsentation in „Step 4“ zwar zu einer sehr schnellen Berechnung, verhindert aber einen direkten Einblick in die Konstruktion. Man muss sich erst mit der Arbeitsweise der Funktion „Filter“ und deren Einstellmöglichkeiten vertraut machen, um zu verstehen welchen Beitrag sie für die Parameterschätzung liefert.

Da wir uns in erster Linie mit dem AR(1)-ARCH(1)-Fall befassen wollen, soll das genaue Zusammenspiel der 6 Steps erst an der analogen Umsetzung des Programmcodes für diesen Fall geschehen.

---

<sup>11</sup><https://svn.r-project.org/Rmetrics/trunk/fGarch/R/garchFit.R>

```

> ar1arch1Fit=function(x)
{
#Step 1: Initialize Time Series Globally:
x<-x

#Step 2: Initialize Model Parameters and Bounds
Mean=mean(x); Var=var(x); S=1e-6
params=c(mu=Mean, alpha0=0.1*Var, alpha1=0.2, beta=0.8)
lowerBounds=c(mu=-10*abs(Mean), alpha0=S^2, alpha1=S, beta=S)
upperBounds=c(mu=10*abs(Mean), alpha0=100*Var, alpha1=1-S, beta=1-S)

#Step 3: Set Conditional Distribution Function:
ar1arch1Dist=function(z, hh){dnorm(x=z/hh)/hh}

#Step 4: Compose log-Likelihood Function:
ar1arch1LLH=function(parm){
mu=parm[1]; alpha0=parm[2]; alpha1=parm[3]; beta=parm[4]
u=1
p=1
N=length(x)
z=rep(0, N)
for(i in 2:N){z[i]=x[i]-mu-(beta*x[i-(1:u)])}
mvar=mean(z^2)
h=rep(alpha0+alpha1*mvar,N)
for(i in 2:N){h[i]=alpha0+sum(alpha1*(abs(z[i-(1:p)]))^2)}
hh=sqrt(abs(h))
llh=-sum(log(ar1arch1Dist(z, hh)))
llh}
print(ar1arch1LLH(params))

#Step 5: Estimate Parameters and Compute Numerically Hessian:
fit=nlminb(start=params, objective=ar1arch1LLH,
lower=lowerBounds, upper=upperBounds, control=list(trace=1))
epsilon=0.0001*fit$par
Hessian=matrix(0, ncol=4, nrow=4)
for (i in 1:4){
for (j in 1:4){
x1=x2=x3=x4=fit$par
x1[i]=x1[i]+epsilon[i]; x1[j]=x1[j]+epsilon[j]
x2[i]=x2[i]+epsilon[i]; x2[j]=x2[j]-epsilon[j]
x3[i]=x3[i]-epsilon[i]; x3[j]=x3[j]+epsilon[j]
x4[i]=x4[i]-epsilon[i]; x4[j]=x4[j]-epsilon[j]
Hessian[i,j]=(ar1arch1LLH(x1)-ar1arch1LLH(x2)-ar1arch1LLH(x3)+ar1arch1LLH(x4))/
(4*epsilon[i]*epsilon[j])
}
}
}

```

```

#Step 6: Create and Print Summary Report:
se.coef=sqrt(diag(solve(Hessian)))
tval=fit$par/se.coef
matcoef=cbind(fit$par, se.coef, tval, 2*(1-pnorm(abs(tval))))
dimnames(matcoef)=list(names(tval), c("Est.", "Std.Err.", "t value", "Pr(>|t|)"))
cat("\nCoefficient(s):\n")
printCoefmat(matcoef, digits=6, signif.stars=TRUE)
}

```

Die Funktionsweise eines solchen Algorithmus zur Parameterschätzung, lässt sich nun anhand der sechs vorgegebenen Schritte sehr schön nachvollziehen. Für eine etwas detailliertere Erläuterung, sei wiederum auf den Anhang (Abschnitt 4.7) verwiesen. Nennenswert an dieser Stelle ist vielleicht der Umstand, dass im obigen Algorithmus ein  $\mu$  auftaucht und so eigentlich ein zusätzlicher Parameter ausgegeben wird. Dieser ist wiederum einer Datentransformation geschuldet und wird von dem in R genutzten Lösungsverfahren `nlminb` explizit verlangt (vgl. [22] - S.5 oder `nlminb`-Hilfe in R). Die Notwendigkeit der Datentransformation wurde oben ja thematisiert, hat aber den Nebeneffekt, dass die resultierenden Daten sich in einem Wertebereich bewegen, der keine inhaltlich sinnvolle Interpretation zulässt. Über ein „Intercept“  $\mu$  (vgl. Ausgabe der Parameterschätzung in R), findet eine Zentrierung am Gesamtmittelwert statt und die Daten werden dahingehend interpretierbar, dass man von einem Plus oder Minus, einem Zuwachs oder Abfall sprechen kann. Vor diesem Hintergrund bietet sich auch ein Blick auf die verschiedenen Realisationen/Simulationen an, die trotz unterschiedlicher Parameterkonstellationen eben alle diese Schwankung um einen zentralen Wert aufweisen.

Auch wenn die einzelnen Optimierungsschritte von `nlminb` übernommen werden, bietet der `ar1arch1Fit`-Algorithmus einen wesentlich besseren Einblick in die Konstruktion und Arbeitsweise, als es `garchFit` kann. Man muss zwar auf die breite Anwendungspalette von `garchFit` verzichten, findet in diesem überschaubaren Programmcode jedoch leichter einen Ansatzpunkt für Erweiterungen oder Abwandlungen.

Die Funktion `Sar1arch1Fit` ist eine Erweiterung der `ar1arch1Fit`-Implementierung und liefert zu vorgegebenen Werten von  $\underline{w}_0$ ,  $\overline{w}_0$ ,  $\alpha$  und  $\gamma$  die Parameterschätzung für den geglätteten AR(1)-ARCH(1)-Prozess (Geradenbeispiel). Die Vorteile eines selbstgeschriebenen und überschaubaren Programms, werden hier sehr deutlich. Bei der Wahl  $\alpha=\gamma=1$ , entsprechen die Ergebnisse natürlich denen von `ar1arch1Fit`. Die Unterschiede in der Konstruktion beschränken sich auf „Step3“ und „Step4“, weshalb der Programmcode auch nur verkürzt dargestellt wird.



```

> Sar1arch1Fit=function(x)
{
  ...
  #Step 3: Set Conditional Distribution Function:
  Sar1arch1Dist=function(z, hh, kk){dnorm(x=z/hh)/kk}

  #Step 4: Compose log-Likelihood Function:
  Sar1arch1LLH=function(parm){
    mu=parm[1]; alpha0=parm[2]; alpha1=parm[3]; beta=parm[4]
    u=1
    p=1
    alpha=0.5
    gamma=0.5
    w0oben=0.12
    w0unten=-0.15
    N=length(x)
    z=y=rep(0, N)
    for(i in 2:N){
      if(x[i]>w0oben){z[i]=(x[i]-w0oben)/gamma+w0oben-mu-(beta*x[i-(1:u)])}
      if(x[i]<w0unten){z[i]=(x[i]-w0unten)/alpha+w0unten-mu-(beta*x[i-(1:u)])}
      if(x[i]<=w0oben & x[i]>=w0unten){z[i]=x[i]-mu-(beta*x[i-(1:u)])}
      y[i]=x[i]-mu-(beta*x[i-(1:u)])}
    mvar=mean(y^2)
    h=rep(alpha0+alpha1*mvar,N)
    k=rep(alpha0+alpha1*mvar,N)
    for(i in 2:N){h[i]=alpha0+sum(alpha1*(abs(y[i-(1:p)]))^2)}
    hh=sqrt(abs(h))
    for(i in 2:N){
      if(x[i]>w0oben){k[i]=(alpha0+sum(alpha1*(abs(y[i-(1:p)]))^2))*((abs(gamma))^2)}
      if(x[i]<w0unten){k[i]=(alpha0+sum(alpha1*(abs(y[i-(1:p)]))^2))*((abs(alpha))^2)}
      if(x[i]<=w0oben & x[i]>=w0unten){k[i]=alpha0+sum(alpha1*(abs(y[i-(1:p)]))^2)}
    }
    kk=sqrt(abs(k))
    llh=-sum(log(Sar1arch1Dist(z, hh, kk)))
    llh}
    print(Sar1arch1LLH(params))

  ...}

```

Die wesentlichen Unterschiede zum `ar1arch1Fit`-Beispiel ergeben sich aus der Tatsache, dass man hervorgerufen durch die Glättungsfunktion keine Normalverteilung mehr betrachtet (vgl. Anwendung des Transformationssatzes für Dichten weiter oben):

$$f_{\hat{g} \circ X}(y) = \begin{cases} f_X(\hat{g}_1^{-1}(y)) \cdot \left| \frac{\partial \hat{g}_1^{-1}(y)}{\partial y} \right| = \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((\frac{y-w_0}{\alpha} + w_0) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\alpha} \right| & \text{für } y < \underline{w_0} \\ f_X(\hat{g}_2^{-1}(y)) \cdot \left| \frac{\partial \hat{g}_2^{-1}(y)}{\partial y} \right| = \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(y - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot |1| & \text{für } \underline{w_0} < y < \overline{w_0} \\ f_X(\hat{g}_3^{-1}(y)) \cdot \left| \frac{\partial \hat{g}_3^{-1}(y)}{\partial y} \right| = \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((\frac{y-\overline{w_0}}{\gamma} + \overline{w_0}) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\gamma} \right| & \text{für } \overline{w_0} < y. \end{cases}$$

Genau diese hier zu Grunde liegende Dichte, wird durch den obigen Algorithmus umgesetzt. Die Prozedur `Sar1arch1Fit` wirkt durch diese Anpassungen zunächst etwas unübersichtlicher, folgt aber genau den gleichen Überlegungen bzw. Berechnungsschritten, wie `arlarch1Fit` im unglätteten Fall.

`arlarch1Fit`-Ausgabe (Log>Returns der Allianzdaten)

```
Coefficient(s):
      Estimate   Std. Error   t value   Pr(>|t|)
mu      -2.54236e-04  2.79508e-04  -0.90958    0.36304
alpha0   2.79252e-04  8.81028e-06  31.69616 < 2.22e-16 ***
alpha1   4.03011e-01  3.34027e-02  12.06522 < 2.22e-16 ***
beta     1.37942e-01  2.08273e-02   6.62312  3.517e-11 ***
--
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Mit der Festlegung  $\underline{w_0} = -0.15$ ,  $\overline{w_0} = 0.12$ ,  $\alpha = 0.5 = \gamma$  (siehe nächster Abschnitt), erhält man im glätteten Fall das Ergebnis:

`Sar1arch1Fit`-Ausgabe (Log>Returns der Allianzdaten)

```
Coefficient(s):
      Estimate   Std. Error   t value   Pr(>|t|)
mu      -2.57695e-04  2.79503e-04  -0.92197    0.35654
alpha0   2.79295e-04  8.81482e-06  31.68476 < 2.22e-16 ***
alpha1   4.04023e-01  3.34522e-02  12.07762 < 2.22e-16 ***
beta     1.38581e-01  2.08253e-02   6.65444    2.8438 ***
--
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Einen ersten Anhaltspunkt zur Beurteilung der geschätzten Parameter, liefern die angegebenen Standardfehler und die Signifikanzskala. Die Tatsache, dass die Standardfehler nahe bei 0 liegen und, insbesondere im Verhältnis zu den Schätzungen selbst, als klein angesehen werden können, ist ein Indiz für die Güte der Parameterschätzung. Des weiteren deuten die drei Sternchen hinter den Parametern  $\alpha_0$ ,  $\alpha_1$  und  $\beta$  an, wie wichtig die einzelnen Parameter für die Anpassung

des Modells an die gegebenen Daten sind (\*\* = sehr hoch).

Um einen analogen Algorithmus für den polynomialen Glättungsansatz zu konstruieren, bedarf es wiederum einer Anwendung des Transformationssatzes. Die Konsequenzen auf die Dichtefunktion, lassen sich nämlich nicht so ohne weiteres ablesen:

$$\bar{g}(w') := g^{4*}(w') = \begin{cases} (\underline{w}_0 + 1) - (\underline{w}_0 + 1 - w')^\alpha & , w' < \underline{w}_0 \\ w' & , \underline{w}_0 \leq w' \leq \overline{w}_0 \\ (\overline{w}_0 - 1) + (w' - \overline{w}_0 + 1)^\gamma & , \overline{w}_0 < w' \end{cases} \quad 0 < \alpha, \gamma < 1$$

Die Ausgangssituation  $X \sim \mathcal{N}(\beta w, \alpha_0 + \alpha_1 w^2)$  bleibt die gleiche und die gesuchte Dichte ist  $f_{\bar{g} \circ X}$ .

Gemäß der Forderung im Transformationssatz und der Konstruktion der Glättungsfunktion, zerlegt man den Definitionsbereich von  $\bar{g}$  in die drei offenen Intervalle

$$I_1 = ] - \infty, \underline{w}_0[, \quad I_2 = ]\underline{w}_0, \overline{w}_0[, \quad I_3 = ]\overline{w}_0, +\infty[.$$

Mit der Einschränkung auf eines dieser drei Intervalle, hat man es jeweils mit einer stetig differenzierbaren Bijektion zu tun und kann dementsprechend die Umkehrfunktion bestimmen.

$$\bar{g}_1 = \bar{g}|_{I_1} \Rightarrow \bar{g}_1^{-1}(y) = \underline{w}_0 + 1 - (\underline{w}_0 + 1 - y)^{\frac{1}{\alpha}}$$

$$\bar{g}_2 = \bar{g}|_{I_2} \Rightarrow \bar{g}_2^{-1}(y) = \text{Id}(y) = y$$

$$\bar{g}_3 = \bar{g}|_{I_3} \Rightarrow \bar{g}_3^{-1}(y) = (y - \overline{w}_0 + 1)^{\frac{1}{\gamma}} + \overline{w}_0 - 1$$

Offenbar ist auch  $\frac{\partial g_m(x)}{\partial x} \neq 0$  für  $m \in \{1, 2, 3\}$  erfüllt und

$$\left| \frac{\partial \bar{g}_1^{-1}(y)}{\partial y} \right| = \left| -\frac{1}{\alpha} (\underline{w}_0 + 1 - y)^{\frac{1}{\alpha} - 1} \cdot (-1) \right| = \frac{1}{\alpha} (\underline{w}_0 + 1 - y)^{\frac{1-\alpha}{\alpha}} > 0,$$

$$\left| \frac{\partial \bar{g}_2^{-1}(y)}{\partial y} \right| = |1| = 1 > 0,$$

$$\left| \frac{\partial \bar{g}_3^{-1}(y)}{\partial y} \right| = \left| \frac{1}{\gamma} (y - \overline{w}_0 + 1)^{\frac{1}{\gamma} - 1} \cdot 1 \right| = \frac{1}{\gamma} (y - \overline{w}_0 + 1)^{\frac{1-\gamma}{\gamma}} > 0.$$

Damit sind alle Voraussetzungen gegeben und der Transformationssatz liefert die Dichte

$$f_{\bar{g} \circ X}(y) = \begin{cases} f_X(\bar{g}_1^{-1}(y)) \cdot \left| \frac{\partial \bar{g}_1^{-1}(y)}{\partial y} \right| = \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(\underline{w}_0 + 1 - (\underline{w}_0 + 1 - y)^{\frac{1}{\alpha}} - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\alpha} (\underline{w}_0 + 1 - y)^{\frac{1-\alpha}{\alpha}} \right| & \text{für } y < \underline{w}_0 \\ f_X(\bar{g}_2^{-1}(y)) \cdot \left| \frac{\partial \bar{g}_2^{-1}(y)}{\partial y} \right| = \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(y - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot |1| & \text{für } \underline{w}_0 < y < \overline{w}_0 \\ f_X(\bar{g}_3^{-1}(y)) \cdot \left| \frac{\partial \bar{g}_3^{-1}(y)}{\partial y} \right| = \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((y - \overline{w}_0 + 1)^{\frac{1}{\gamma}} + \overline{w}_0 - 1 - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\gamma} (y - \overline{w}_0 + 1)^{\frac{1-\gamma}{\gamma}} \right| & \text{für } \overline{w}_0 < y. \end{cases}$$

Mit diesem Ergebnis lässt sich der Algorithmus für den polynomialen Fall ansetzen, indem man `Sar1arch1Fit` entsprechend modifiziert:

```
> PolySar1arch1Fit=function(x)
{
  ...
  #Step 3: Set Conditional Distribution Function:
  PolySar1arch1Dist=function(z, hh, kk){dnorm(x=z/hh)/kk}

  #Step 4: Compose log-Likelihood Function:
  PolySar1arch1LLH=function(parm){
    mu=parm[1]; alpha0=parm[2]; alpha1=parm[3]; beta=parm[4]
    u=1
    p=1
    alpha=0.5
    gamma=0.5
    w0oben=0.12
    w0unten=-0.15
    N=length(x)
    z=y=rep(0, N)
    for(i in 2:N){
      if(x[i]>w0oben){z[i]=(x[i]-w0oben+1)^(1/gamma)+w0oben-1-mu-(beta*x[i-(1:u)])}
      if(x[i]<w0unten){z[i]=w0unten+1-(w0unten+1-x[i])^(1/alpha)-mu-(beta*x[i-(1:u)])}
      if(x[i]<=w0oben & x[i]>=w0unten){z[i]=x[i]-mu-(beta*x[i-(1:u)])}
      y[i]=x[i]-mu-(beta*x[i-(1:u)])}
    mvar=mean(y^2)
    h=rep(alpha0+alpha1*mvar,N)
    k=rep(alpha0+alpha1*mvar,N)
    for(i in 2:N){h[i]=alpha0+sum(alpha1*(abs(y[i-(1:p)]))^2)}
    hh=sqrt(abs(h))
    for(i in 2:N){
      if(x[i]>w0oben){k[i]=(alpha0+sum(alpha1*(abs(y[i-(1:p)]))^2))*((abs(gamma))^2)*
        (x[i]-w0oben+1)^(2*gamma/(1-gamma))}
      if(x[i]<w0unten){k[i]=(alpha0+sum(alpha1*(abs(y[i-(1:p)]))^2))*((abs(alpha))^2)*
        (w0unten+1-x[i])^(2*alpha/(1-alpha))}
      if(x[i]<=w0oben & x[i]>=w0unten){k[i]=alpha0+sum(alpha1*(abs(y[i-(1:p)]))^2)}
    }
    kk=sqrt(abs(k))
    llh=-sum(log(PolySar1arch1Dist(z, hh, kk)))
    llh}
    print(PolySar1arch1LLH(params))

  ...}
}
```

Mit der Festlegung wie im Geradenbeispiel  $\underline{w}_0 = -0.15$ ,  $\overline{w}_0 = 0.12$ ,  $\alpha = 0.5 = \gamma$ , erhält man im polynomialen Fall das Ergebnis:

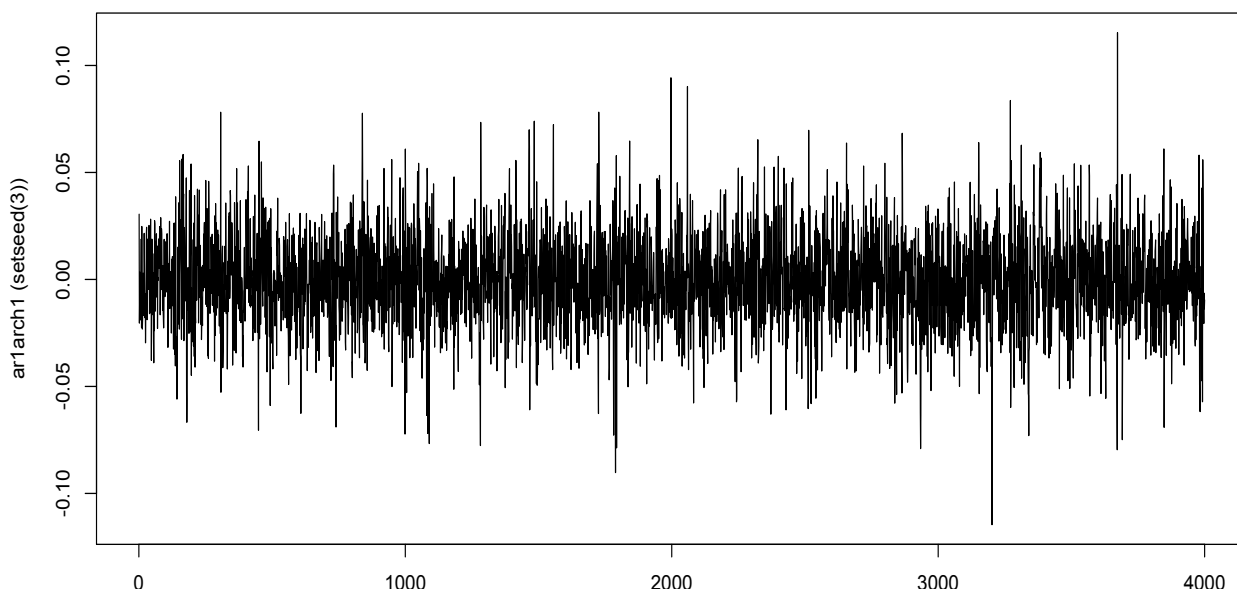
## PolySar1arch1Fit-Ausgabe

```
Coefficient(s):
      Estimate Std. Error  t value    Pr(>|t|)
mu      -2.57702e-04  2.79490e-04  -0.92205    0.3565
alpha0   2.79295e-04  8.81479e-06  31.68478 < 2.22e-16 ***
alpha1   4.04030e-01  3.34525e-02  12.07773 < 2.22e-16 ***
beta     1.38583e-01  2.08252e-02   6.65457    2.8413 ***
--
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Die Resultate unterscheiden sich nicht besonders stark von denen des Geradenbeispiels, was allerdings wenig verwunderlich erscheint, wenn man sich an die Konstruktion beider Ansätze erinnert. So müssen sich die auftauchenden Werte im Datenmaterial erst einmal weit von dem angesetzten Kompaktum  $[w_0, \overline{w_0}]$  entfernen, um einen gravierenden Effekt hervorzurufen. Da dies hier nur für einen Bruchteil der Datenpunkte (vgl. nächster Abschnitt) der Fall ist, sind die ähnlichen Ergebnisse nachvollziehbar.

### 3.4.1 Güte der Parameterschätzung

Für eine genaue Beurteilung der Parameterschätzung bzw. um die Ergebnisse greifbar und vor allem anschaulich zu machen, kommt man um einen grafischen Vergleich nicht herum. Simuliert man mit den geschätzten Parametern einen AR(1)-ARCH(1)-Prozess und vergleicht das Schaubild mit den Log>Returns der Allianz-Daten, so fallen einem doch größere Unterschiede auf. Man könnte sogar den Eindruck gewinnen, dass diese AR(1)-ARCH(1)-Simulation herzlich wenig mit dem Allianzdatensatz gemein hat.



Bedenkt man jedoch, dass die Allianz-Daten (Log>Returns) nur als eine einzige Realisation eines stochastischen Prozesses aufgefasst werden, so wäre es eher verwunderlich, wenn sich

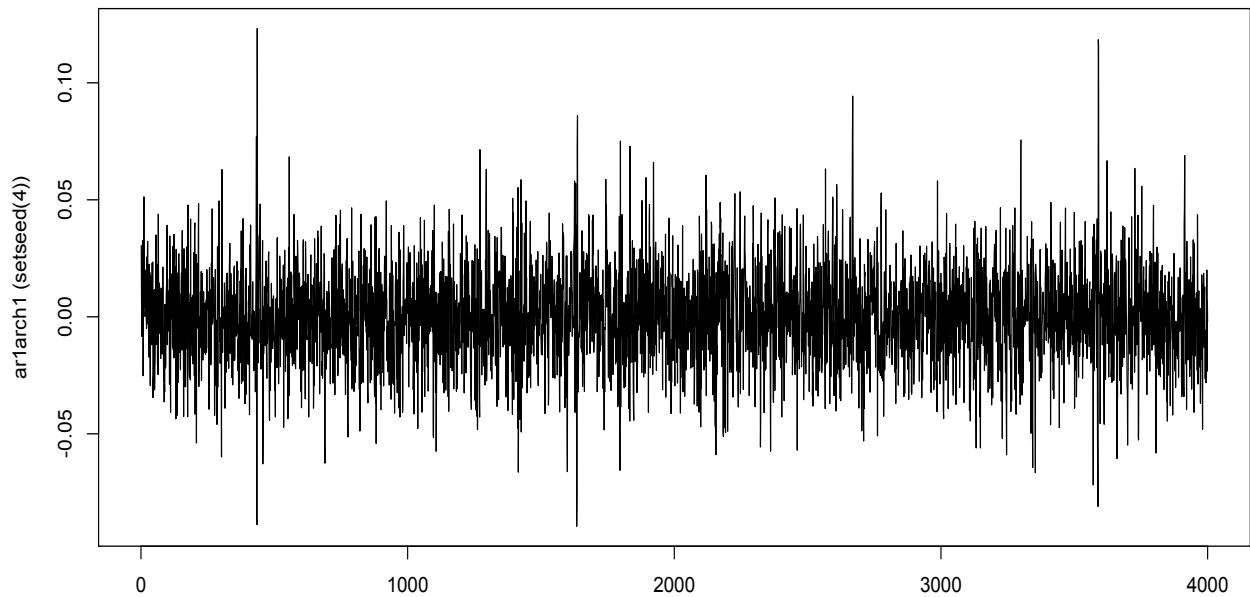


Abbildung 3.7: AR(1)-ARCH(1)-Simulationen mit den Ergebnissen der Parameterschätzung. Nur der Pfad des Störterms unterscheidet sich.

die Log>Returns und die Simulation zu sehr entsprechen. Vielmehr hängt das Schaubild einer Simulation sehr stark davon ab, wie der White-Noise Prozess gestartet wird. Mit verschiedenen Startwerten für den Zufallsgenerator<sup>12</sup>, lassen sich unterschiedlichste Variationen erzeugen. Diese Variationen können sich durchaus in zentralen Punkten unterscheiden, wie etwa die Position und Höhe von Ausreißern deutlich macht - vgl. Abb. 3.7.

Die geschätzten Parameterwerte liefern demnach keine Anleitung zur Rekonstruktion der Allianz-Daten. Entsprechend dem Maximum-Likelihood-Prinzip sind die Ergebnisse so zu interpretieren, dass die vorliegenden Daten mit größter Wahrscheinlichkeit aus einem solchen Modell stammen, das diese Parameterwerte hat. Das Modell und die Parameterzahl werden hierbei als Hypothese zugrunde gelegt, so dass die Aussage nur als Wertung innerhalb dieses Rahmens gesehen werden kann. Mit mehr Parametern oder einem völlig anderen Modell, mag eine bessere Anpassung möglich sein. Unsere Zielsetzung soll es jedoch nicht sein, die beste AR(u)-ARCH(p) Anpassung vorzunehmen. Vielmehr soll das Augenmerk darauf liegen, ob man die Anpassung eines als „gut“ empfundenen Modells nicht durch eine überlagerte Glättungsfunktion noch verbessern kann.

Ein weiteres Indiz für die Qualität der hervorgebrachten Ergebnisse, kann die Rücktransformation zu den Ausgangsdaten liefern. Sollten sich dabei keine Ähnlichkeiten in den Schaubildern zeigen, so wäre sicherlich von einer Fehlanpassung auszugehen.

Wie bereits der Vergleich von Log>Returns und AR(1)-ARCH(1)-Simulation nahelegte, ist eine 100%ige Übereinstimmung mit dem ursprünglichen Allianz-Kurs, natürlich nicht gelungen. Allein das Spektrum, in dem die Werte nach der Rücktransformation angenommen werden (der Startwert ist identisch), unterscheidet sich im ersten Fall gravierend. Abgesehen davon, erkennt man aber unschwer einige gemeinsame Charakteristika.

<sup>12</sup>`set.seed` (siehe White-Noise im Anhang - Abschnitt 4.5)

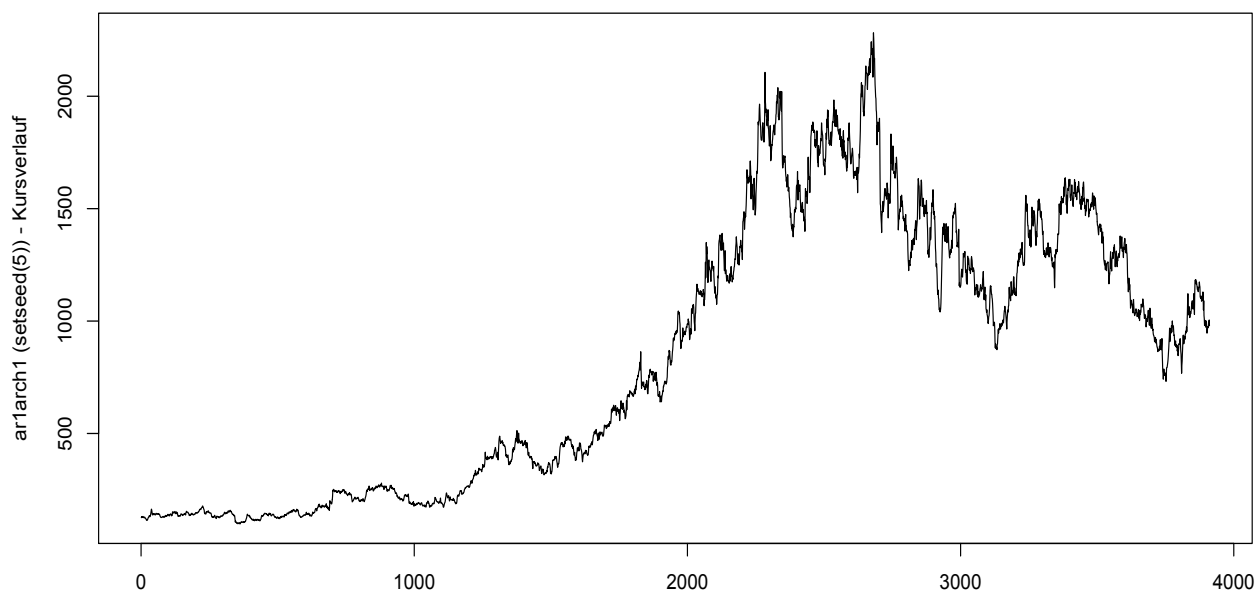


Abbildung 3.8: Rücktransformation einer AR(1)-ARCH(1)-Simulation.

So findet in beiden Fällen ein von Unterbrechungen durchzogener Anstieg statt, der im Bereich  $t \in [2000, 3000]$  seinen Höhepunkt findet. Man könnte auch meinen, die 2 Höchstwerte in diesem Bereich würden, zumindest von der Position her, denen in den Allianz-Daten ziemlich genau entsprechen. Der im Aktienkurs um etwa  $t=3000$  beginnende Verfall, der die Werte wieder auf ein viel niedrigeres Niveau bringt, ist hier aber nicht erkennbar. Dies verwundert jedoch weniger, da diese Trendwende in einem als stationär geltenden Modell, ohne äußere Eingriffe nicht zu erwarten ist (Strukturbruch - tiefgreifende Veränderung des Gesamtbildes). So zeichnen sich, bei einem betrachteten Zeitfenster von 02.01.1990 - 30.12.2004, im Kursverlauf der Allianz-Aktie, deutschland- oder auch weltweite Phänomene ab. Viele Kurseinbrüche oder auch Trendwenden im Kursverlauf, sind den äußeren Umständen geschuldet. Man denke nur an Wirtschaftskrisen, Kriege und auch positive Effekte, wie den wirtschaftlichen Aufschwung. Nennenswerte Beispiele in diesem Zusammenhang wären: Golf-Krieg (1990), Wirtschaftskrise in Süd-Ost Asien (1997), Finanzkrise in Russland (1998) und die Angriffe auf das World Trade Center (2001).

Hält man sich vor Augen, dass die geschätzten Parameter zusammen mit einer White-Noise-Störgröße für diese Simulation verantwortlich zeichnen, so erscheint es bei den erkennbaren Ähnlichkeiten durchaus nachvollziehbar, dass ein anderer Pfad des White-Noise-Prozesses ein noch passenderes Bild erzeugen könnte. Andererseits kann der White-Noise-Prozess aber auch ein völlig verändertes Schaubild hervorrufen, dem man keine Ähnlichkeit mit den Ausgangsdaten mehr attestieren würde. Die folgende Simulation ist aus der gleichen Parameterkonstellation hervorgegangen und nur ein veränderter Startwert für den Zufallsgenerator, der ja dem White-Noise Prozess zu Grunde liegt, bringt dieses völlig anders wirkende Ergebnis.

Der Wertebereich  $[50, 350]$  passt in diesem Fall zwar viel besser zu dem, durch die Allianz-Daten vorgegebenem Spektrum, jedoch findet man im Kurvenverlauf nicht mehr die gleichen Charakteristika wie oben. Dies ist also ein eindrucksvolles Beispiel für die Variationsfähigkeit

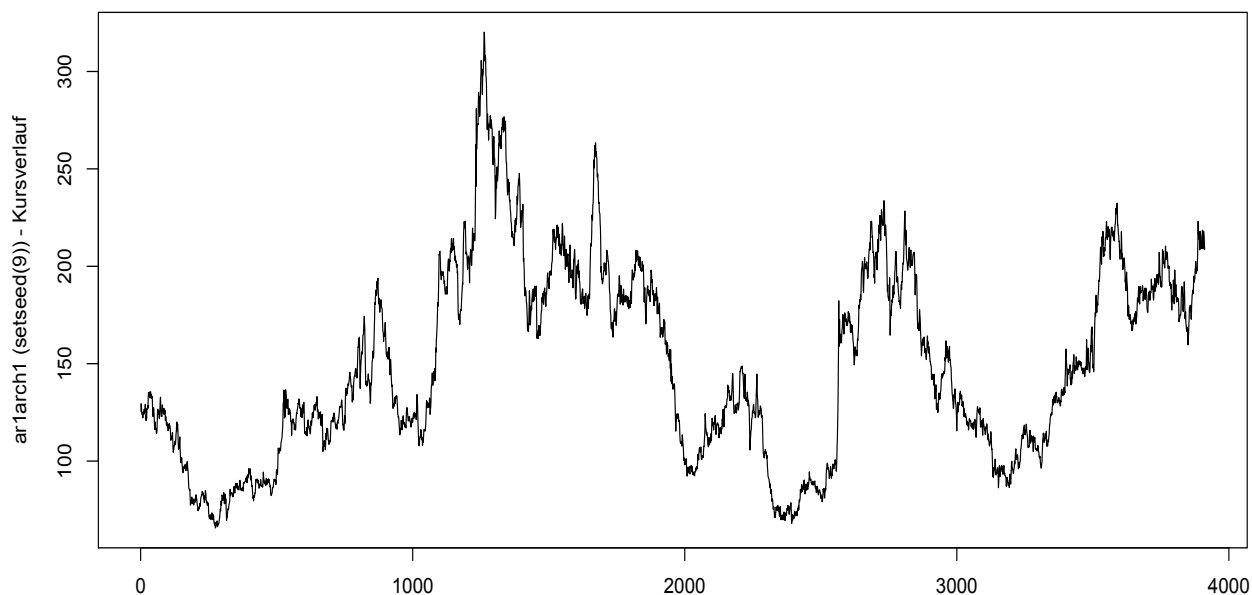


Abbildung 3.9: Rücktransformation einer AR(1)-ARCH(1)-Simulation.

des AR(1)-ARCH(1)-Modells und stellt noch mal heraus, dass jede Simulation nur eine einzige Realisation ist.

Auf der anderen Seite liefern die nur durch veränderte White-Noise Störgrößen hervorgerufenen Ergebnisse, die Grundlage für eine handfeste Beurteilung der Parameteranpassung.

Letztendlich muss man nur die Blickrichtung ändern: Für die obigen Simulationen wurde eine Realisation eines White-Noise Prozesses vorgegeben und auf dieser Grundlage dann ein möglicher Kurvenverlauf eines AR(1)-ARCH(1)-Modells errechnet. Von diesem Standpunkt aus gesehen, wäre es also nur reiner Zufall, wenn man genau den gewünschten Kurvenverlauf generiert hätte.

Nimmt man jedoch die geschätzten Parameter und den AR(1)-ARCH(1)-Modellrahmen als passend an, so kann man auch den umgekehrten Weg beschreiten. Aus den Allianz-Daten oder anders ausgedrückt aus der „angenommenen“ Realisation eines AR(1)-ARCH(1)-Prozesses, lassen sich die  $\varepsilon_t$  (Störgrößen) herausrechnen und bilden letztendlich eine eigene Zeitreihe. Das Auflösen nach diesen zufallsbedingten Störgrößen und deren weitere Untersuchung, wird auch mit dem Begriff Residuenanalyse bezeichnet. Bei einer gelungenen Modellanpassung sollten die verbliebenen Störgrößen also Zufallscharakter besitzen oder wie in diesem Fall einem White-Noise Prozess folgen.

In der klassischen Zeitreihenanalyse wird die Zeitreihe in drei Komponenten zerlegt: Trend  $T$ , Saison  $S$  und Residuen  $e$  - vgl. z.B. [39] S.7, [10] oder Schlittgen [27] S.17ff. Dabei bezeichnet der Trend die langfristige Entwicklung, die Saisonkomponente soll wiederkehrende Muster beschreiben (meist jährlich) und die Residuen fassen zufallsabhängige Schwankungen zusammen.

Im **additiven Modell** vermutet man einen Aufbau von der Art

$$x_t = T_t + S_t + e_t, \quad t = 1, 2, \dots, n$$



und im **multiplikativen Modell**

$$x_t = T_t \cdot S_t \cdot e_t, \quad t = 1, 2, \dots, n.$$

Hat man die Trend- und Saisonkomponente bestimmt, so kann man nach den Residuen auflösen. Im additiven Modell bedeutet dies:

$$e_t = x_t - T_t - S_t, \quad t = 1, 2, \dots, n.$$

In der Situation eines AR(1)-ARCH(1)-Modells, ergeben sich die Residuen als:

$$e_t = x_{t+1} - \mu - \beta \cdot x_t \quad \text{für } t \in \{1, \dots, 3910\}.$$

Diese Vorgehensweise entspricht genau dem Prinzip einer klassischen Zeitreihenanalyse (siehe z.B. Brockwell [6] S.34) und ein Vergleich der Ergebnisse, der selbstgeschriebenen Prozeduren<sup>13</sup> zur Residuenanalyse, mit der `garchFit`-Prozedur

```
ar1arch1<-garchFit(~arma(1,0)+garch(1,0),data=diff(log(Allianz)))  
plot(residuals(ar1arch1))
```

liefert völlige Übereinstimmung.

Ein konkretes Beispiel für die Umsetzung auf den AR(1)-ARCH(1)-Fall, findet man in McNeil und Frey [8]. Im Smooth-AR(1)-ARCH(1)-Fall stellt sich die Situation zwar ganz ähnlich dar, aber die Aufteilung in drei Teilgebiete macht die Sache etwas unübersichtlicher. Im Gegensatz zu den obigen zwei Befehlszeilen, erhält man aber eine genaue Vorstellung von den gemachten Rechenschritten.

Da wir `garchFit` nur zum Vergleich nutzen, beruhen auch alle Grafiken auf selbstgeschriebenen Umsetzungen der obigen Definition für Residuen - vgl. Abschnitt 4.8. Festzuhalten bleibt aber, dass sich mittels einer Residuenanalyse kein nennenswerter Unterschied ergibt, zwischen der AR(1)-ARCH(1) und Smooth-AR(1)-ARCH(1)-Anpassung.

Man startet hier mit dem „Ergebnis“ und will daraus die zugrunde liegenden White-Noise Störgrößen bestimmen. Ein Residuendiagramm sollte demnach keine erkennbaren Gesetzmäßigkeiten aufweisen, sondern völlig zufällig wirken.

Die Veranschaulichung liefert zunächst den Eindruck einer zufälligen Verteilung auf positive und negative Werte. Andererseits fallen zwar vereinzelte, aber dafür starke Ausreißer auf. Des Weiteren gibt es offenbar auch einen Unterschied in der ersten und zweiten Hälfte der Datenpunkte. Für  $t \in [0, 2000]$  ist die Streuung geringer als für  $t \in [2000, 4000]$ . Es gibt also durchaus Charakteristika eines White-Noise-Prozesses, aber ebenso Punkte die klar dagegen sprechen. Durch Verbindung der Datenpunkte zu einer typischen Zeitreihengrafik, könnte man den einen oder anderen charakteristischen Punkt noch weiter hervorheben, aber eine greifbare Kennzahl erscheint für die weitere Beurteilung an dieser Stelle wichtiger.

---

<sup>13</sup>siehe Anhang

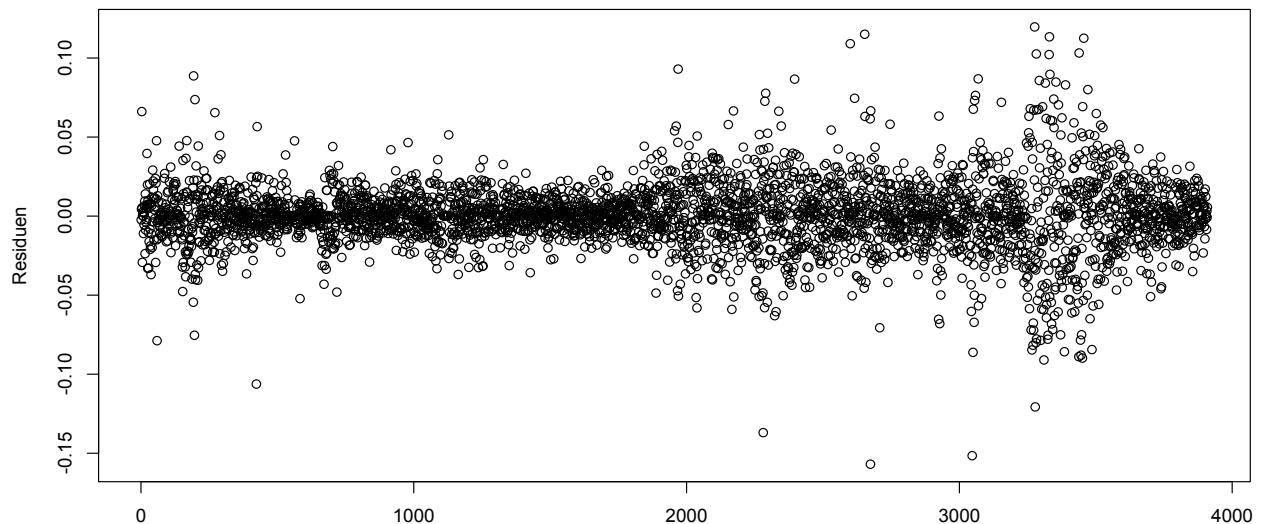


Abbildung 3.10: Residuen der AR(1)-ARCH(1)-Parameterschätzung.

Neben der grafischen Begutachtung, kann ein Test auf Normalverteilung ein etwas handfesteres Kriterium darstellen, um seine Ergebnisse zu verifizieren. Ein in R bereits implementiertes Verfahren, ist der **Shapiro-Wilk-Test** auf Normalverteilung.

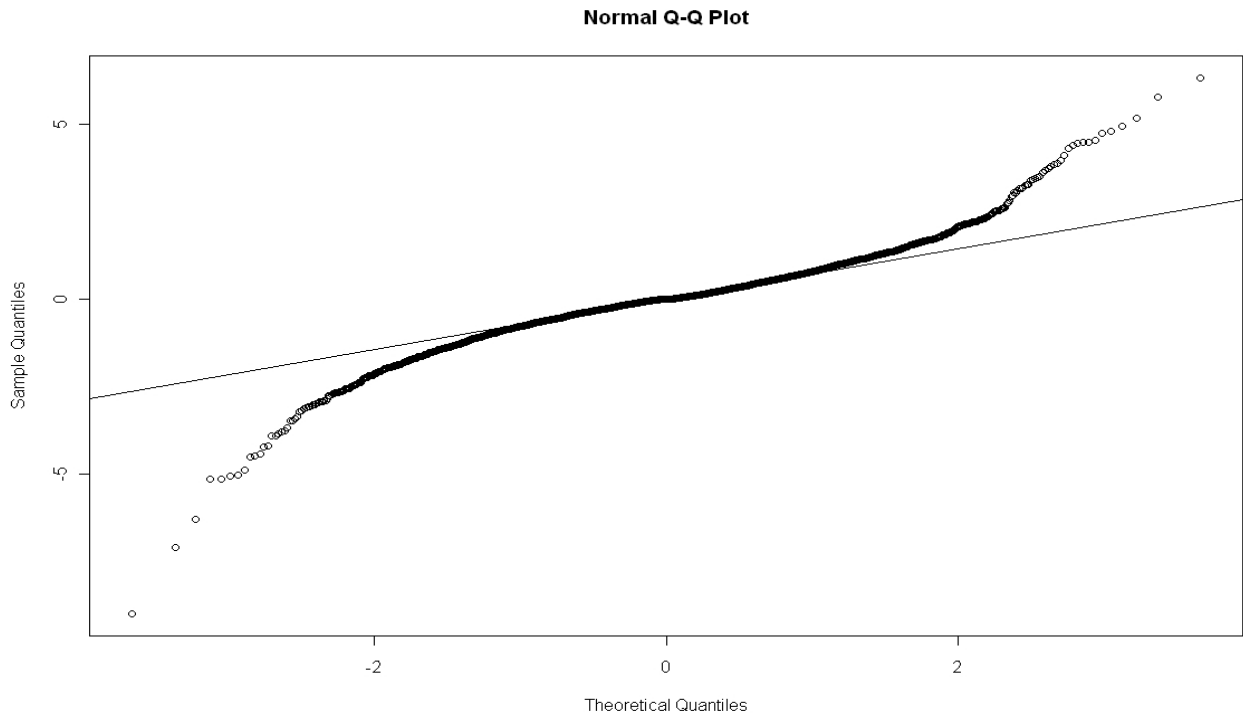
#### Shapiro-Wilk normality test

```
data:    Residuen
W = 0.9243, p-value < 2.2e-16
```

Der ausgegebene P-Wert (p-value) ist zu interpretieren als das kleinste Signifikanzniveau, für das die Nullhypothese  $H_0$  ( $H_0$ : Daten entstammen einer Normalverteilung) bei den vorliegenden Daten gerade noch verworfen wird. In der Literatur zu Statistischen Tests, findet man die klassischen Niveauvorgaben  $\alpha = 0.1$ ,  $\alpha = 0.05$  oder  $\alpha = 0.01$ , die durch den P-Wert alle unterboten werden. D.h. die Nullhypothese wird abgelehnt, stattdessen die Alternative gewählt und zwar sehr deutlich.

Die Residuenanalyse wird also auch durch den Test auf Normalverteilung untermauert und das AR(1)-ARCH(1)-Modell scheint nicht geeignet zu sein, um den Kurvenverlauf der Allianz hinreichend gut beschreiben zu können.

Dieses Ergebnis spiegelt sich auch in einem weiteren optischen Hilfsmittel wider. In einem Normal-Quantil Diagramm werden die empirischen Quantile der Residuen und die theoretischen Quantile einer Standardnormalverteilung verglichen.



Wenn die Beobachtungen tatsächlich aus einer Normalverteilung stammen sollten, dann würde man nur geringfügige Abweichungen von der eingezeichneten Geraden erwarten. Dies ist hier insbesondere für große Werte (dem Betrag nach) jedoch nicht der Fall. Eine ähnliche Situation und identische Schlussfolgerungen, findet man in McNeil und Frey [8], wobei diese Arbeit auch als generelle Quelle zu einer konkreten Parameteranpassung für Finanzmarktdaten herangezogen werden kann.

Zusammenfassend lassen sowohl die grafischen Darstellungen als auch der durchgeführte Test nur den Schluss zu, dass die Anpassung nicht unbedingt als gelungen bezeichnet werden kann. Die Parameterschätzung als solche ist geglückt, wie zum Beispiel die Standardfehler oder wichtige Charakteristika in den Simulationen nahe legen (vgl. `ar1arch1Fit`-Ausgabe). Wäre der Allianz-Kurs jedoch tatsächlich mit einem  $AR(1)$ - $ARCH(1)$ -Modell hinreichend darstellbar, so dürften die Residuen keinen Grund zur Ablehnung der Normalverteilungsannahme liefern.

Prinzipiell gibt es zwei Möglichkeiten dieser Problematik zu begegnen. Entweder man begibt sich auf die Suche nach einem passenderen Modell, um den Kurvenverlauf zu erklären oder man versucht es mit einer Einschränkung des Betrachtungszeitraums. Da die Residuen auf eine Zweiteilung des Datensatzes hindeuten (s.o.), wäre dies sicherlich ein vielversprechender Ansatzpunkt. Wie bereits erwähnt, interessieren wir uns jedoch nicht unbedingt für das bestmögliche Modell zur Beschreibung der Allianz-Daten. Vielmehr stellt sich die Frage, ob es überhaupt zu einer Verbesserung durch Anwendung einer Glättungsfunktion kommt. Da die Unterschiede der Parameterschätzung im  $AR(1)$ - $ARCH(1)$ ,  $Smooth-AR(1)$ - $ARCH(1)$  oder auch  $PolySmooth-AR(1)$ - $ARCH(1)$ -Fall sehr gering sind, kommen die obigen Analyseschritte für einen geglätteten Ansatz zu fast identischen Ergebnissen. Auf diese Art und Weise lässt sich kein Ansatz, als der bessere identifizieren. Dass es für die konkrete Datensituation wesentlich bessere Modelle ge-

ben kann, legen die bisherigen Ergebnisse ja bereits offen. Vielversprechende Kandidaten wären zum Beispiel ARCH(p) oder GARCH(p,q)-Modelle, da diese die auftauchenden Charakteristika abdecken (vgl. z.B. [28]).

Ein Vergleich zwischen der empirischen Verteilung der Allianz-Daten und dem angepassten Prozess, bietet nicht nur die Möglichkeit, die Parameterschätzung zu visualisieren, sondern auch den Zusammenhang zwischen verschiedenen Realisationen deutlicher zu machen. In einem ersten Schritt kann man eine gute Vorstellung von den zu Grunde liegenden Daten erhalten, indem man ein Histogramm zeichnet. Hierbei wird ein „echtes“ Histogramm gewählt (Fläche eines Balkens ist identisch mit der relativen Häufigkeit der in der zugehörigen Klasse liegenden Werte), da sich hieraus auch eine empirische Dichte ableiten lässt. Zur Erzeugung der empirischen Dichte, wird der in R bereits implementierte Kern-Dichteschätzer (vgl. [10] S.69,117) genutzt und liefert die „durchgezogene“ Linie in der Grafik (s.u.). Man spricht dabei auch von einer nicht-parametrischen Dichteschätzung, da keine besondere Klasse von zulässigen Dichten vorgegeben ist, sondern einzig und allein aus den Daten eine Dichte geschätzt wird<sup>14</sup>.

**Definition 3.1. ([10] S.117)**

Sei  $h > 0$  eine feste Zahl, die sogenannte Bandbreite. Weiter sei  $K(\cdot)$  eine gegebene Funktion mit  $K(u) \geq 0$  für alle  $u \in \mathbb{R}$  und  $\int_{-\infty}^{\infty} K(u)du = 1$ , der sogenannte **Kern**. Dann heißt die Funktion

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x - x_i}{h}\right)$$

der **Kern-Dichteschätzer** (mit Bandbreite  $h$  zum Kern  $K$ ) der Häufigkeitsverteilung der Beobachtungen  $x_1, \dots, x_n$ .

Keine erschöpfende Auswahl, aber typische Beispiele sind der **Gauß-Kern**

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right), \quad u \in \mathbb{R},$$

der **Epanechnikov-Kern**

$$K(u) = \begin{cases} \frac{3}{4}(1 - u^2) & -1 \leq u < 1 \\ 0 & \text{sonst} \end{cases}$$

oder der **Rechteckkern**

$$K(u) = \begin{cases} \frac{1}{2} & -1 \leq u < 1 \\ 0 & \text{sonst.} \end{cases}$$

Für weitere Varianten, einen tieferen Einblick in die Kern-Dichteschätzung als solche und auch in die Umsetzung in R, sei verwiesen auf Wang [36].

---

<sup>14</sup>Bei einer parametrischen Dichteschätzung, wäre die Form der Dichte im Prinzip vorgegeben und nur die Parameter zu schätzen.

Das Erscheinungsbild eines Histogramms kann sehr stark von der gewählten Klassenbreite abhängen, weshalb für eine bessere Vergleichbarkeit nicht unbedingt die automatische Vorgabe gewählt werden sollte. Im vorliegenden Fall zeigt sich, dass die Allianz-Daten (Log>Returns) und der AR(1)-ARCH(1)-Prozess nur Werte im Bereich  $[-0.2, 0.2]$  annehmen und die Klassenbreite von daher als 0.01 gewählt werden kann.

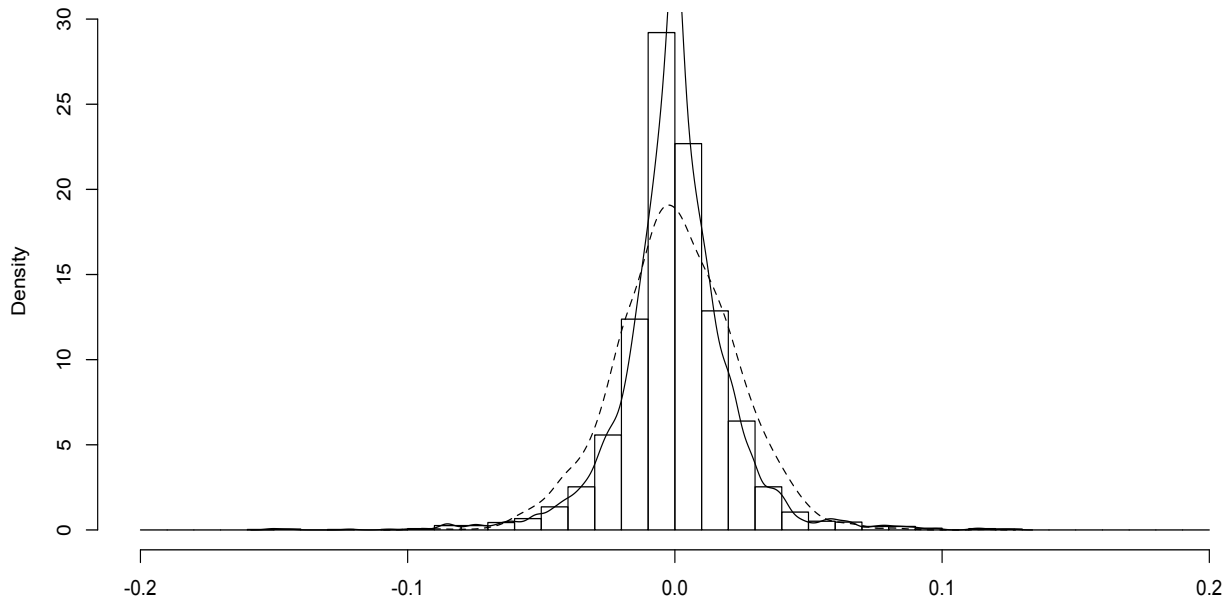


Abbildung 3.11: Histogramm und geschätzte Dichte der Log>Returns zum Allianz Kurs (durchgezogene Linie). Geschätzte Dichte einer AR(1)-ARCH(1)-Realisation (gestrichelte Linie).

Die zusätzlich eingezeichnete „gestrichelte“ Kurve, stellt die empirische Dichte einer AR(1)-ARCH(1)-Realisation mit der Parameterkonstellation aus obiger Schätzung dar. Dazu wird, unter Verwendung der aus der Parameterschätzung gewonnenen Parameter, eine Realisation generiert und die auftretenden Werte nicht nach der Reihenfolge ihres Erscheinens, sondern nur nach ihrer Häufigkeit sortiert. Die Häufigkeitsverteilung auf die einzelnen Klassen, ist dann die Grundlage für die Dichteschätzung. Diese ähnelt auf den ersten Blick einer Normalverteilung, was darin begründet ist, dass die Standardnormalverteilung dem AR(1)-ARCH(1)-Prozess als Störterm dient. Dass es sich dabei aber keineswegs um eine Normalverteilung handelt, merkt man insbesondere an der Abbildung 3.12 (Ausschnitt zu Abbildung 3.11). Aus dieser Gegenüberstellung kann man Rückschlüsse ziehen auf die Beziehung der empirischen Daten zu dem angepassten Prozess und zwar unabhängig von der Reihenfolge der Realisationen. Es spielt demnach keine Rolle, wann ein großer oder kleiner Wert angenommen wurde, sondern nur wie häufig. Man darf nicht außer acht lassen, dass nur eine einzige Realisation betrachtet wird, doch augenfällig werden sehr kleine Werte im Bereich  $[-0.015, 0.015]$  wesentlich häufiger angenommen, als man es für eine AR(1)-ARCH(1)-Realisation (gestrichelte Linie) vermuten würde. Anders ausgedrückt: die empirische Dichte der AR(1)-ARCH(1)-Realisation liegt in diesem Bereich unterhalb der empirischen Dichte der Allianz-Daten. Umgekehrt verhält es sich dann zunächst außerhalb dieses Bereichs. Ein genauerer Blick bzw. eine Anpassung der Skala eröffnet aber noch weitere Unterschiede. So liegt die empirische Dichte der AR(1)-ARCH(1)-Realisation im Bereich der betragsmäßig größten Werte dann wieder unterhalb der empirischen Dichte der

Allianz-Daten. D.h. die dem Betrag nach besonders kleinen und besonders großen Werte, werden anscheinend nicht hinreichend gut erklärt durch die Anpassung.

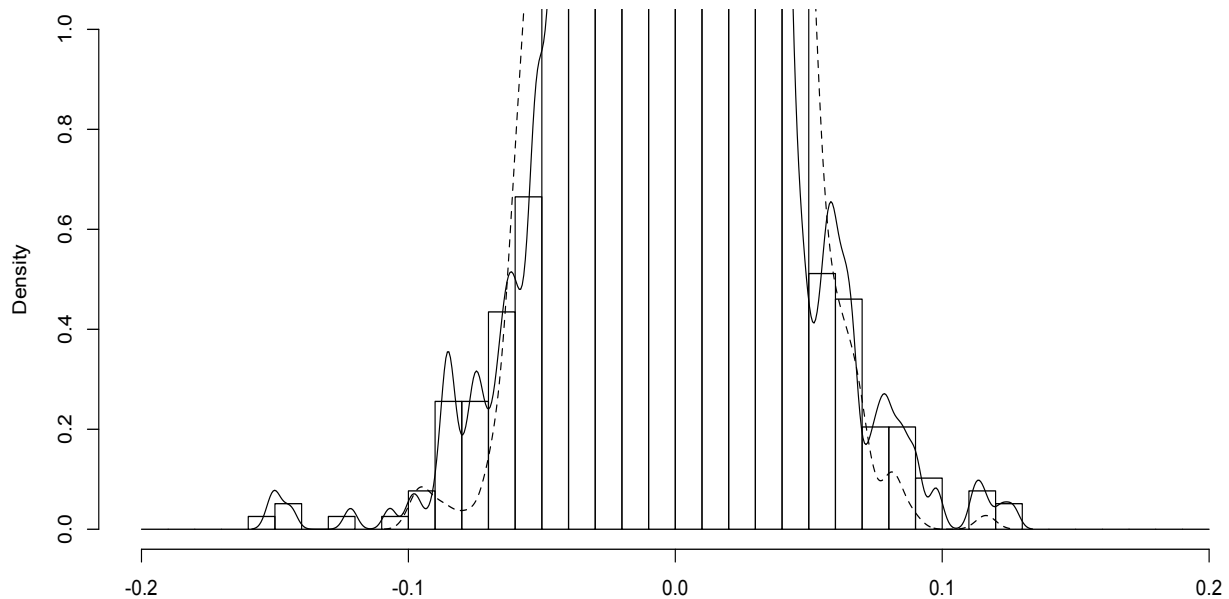


Abbildung 3.12: Ausschnitt zu Abbildung 3.11

Im Anhang (Abschnitt 4.3) wird die Auswirkung der Glättungsfunktion auf die Dichte anhand der typischen Glockenkurve der Normalverteilung untersucht (inkl. grafischer Veranschaulichung). Es ergibt sich, dass die Wahrscheinlichkeitsmasse von rechts in Richtung  $\overline{w}_0$  und von links in Richtung  $\underline{w}_0$  verschoben wird. Das Ausmaß dieser Verschiebung wird durch die Glättungsparameter  $\alpha, \gamma$  gesteuert. Die obige Grafik kann also dahingehend genutzt werden, um eine erste intuitive Wahl von  $w_0, \overline{w}_0, \alpha$  und  $\gamma$  zu treffen. Durch eine geschickte Anpassung der Parameter  $\overline{w}_0$  und  $\underline{w}_0$ , an die vom Betrag her größten beobachteten Werte, kann man die Wahrscheinlichkeit für das Auftreten derartiger Werte im geglätteten Prozess erhöhen oder anders ausgedrückt, die Dichte in diesem Bereich anheben. Bevor man diese Idee jedoch weiter verfolgt, sollte man zunächst einen Eindruck davon gewinnen, ob diese Unterschätzung am Rand nur an der speziellen Realisation des AR(1)-ARCH(1)-Prozesses liegt, oder ein genereller Umstand ist. Dazu werden 500 Realisationen betrachtet und jeweils deren empirische Dichte in obige Zeichnung mit eingebunden. Diesen Realisationen liegen die gleichen Parameter zu Grunde, einzig und allein der Startwert für den Zufallsgenerator des Störterms verändert sich. Wie oben bereits gesehen, können die dadurch erzeugten Schaubilder sich gravierend unterscheiden, obwohl immer die gleiche Gesetzmässigkeit Anwendung findet. Die Zusammengehörigkeit zu ein und der selben Parameterkonstellation, erkennt man in dieser Form jedoch schon eher.

Je nachdem wie der Störterm gestartet wird, kann also auch ohne weitere Modellanpassungen der Ausschlag im Bereich  $[-0.01, 0.01]$  stärker ausfallen, jedoch überragen die beobachteten Ergebnisse diesen Ausschlag in allen 500 Realisationen immer noch sehr deutlich. Für die Beurteilung der Ergebnisse am Rand des angenommenen Wertebereichs, bedarf es wiederum einer Skalenanpassung.

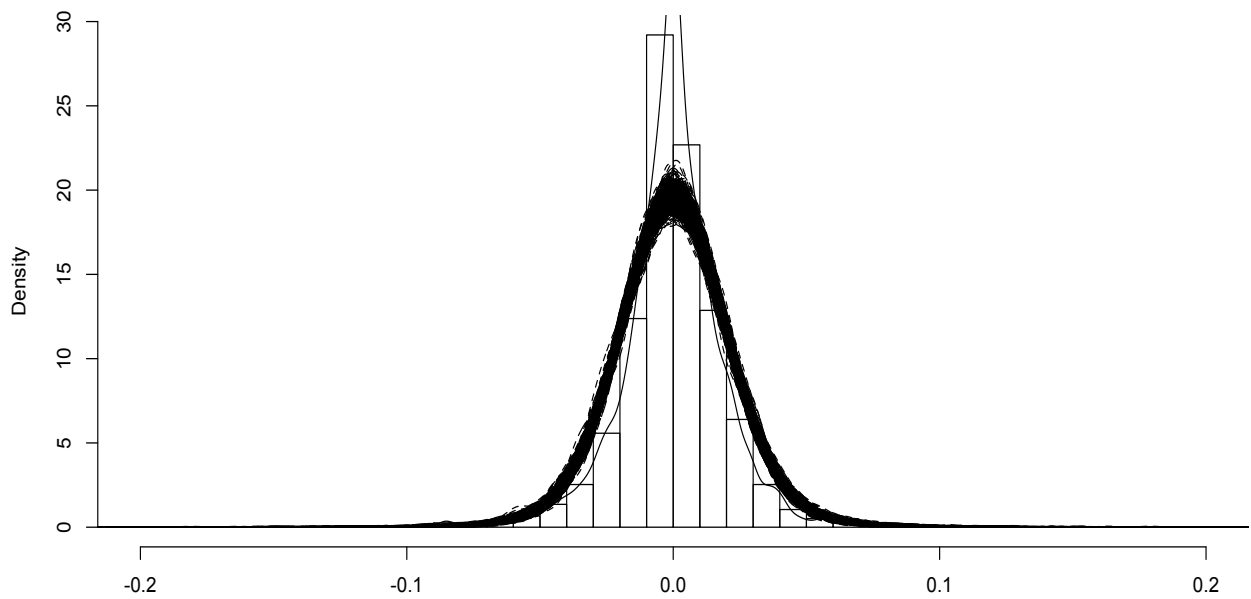


Abbildung 3.13: 500 geschätzte Dichten, zu 500 Realisationen eines AR(1)-ARCH(1).

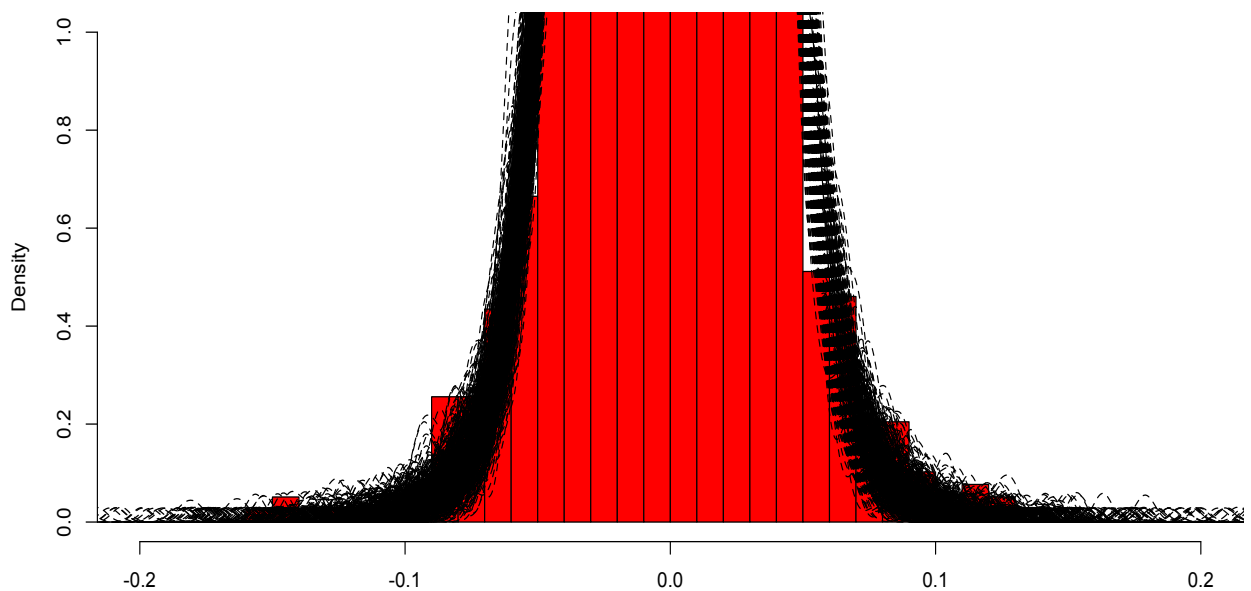


Abbildung 3.14: Ausschnitt zu Abbildung 3.13

Die rote Färbung des Histogramms, dient einzig und allein der besseren Abgrenzung von den 500 Dichte-Kurven und zeigt wie an mehreren Stellen die Ausschläge gar nicht oder nur vereinzelt überdeckt werden.

Es entsteht also der Eindruck, dass im Durchschnitt die Häufigkeit, mit der dem Betrag nach „große“ Werte angenommen werden, durch die AR(1)-ARCH(1)-Anpassung womöglich nicht hinreichend erfasst wird.

Dieser Idee folgend, soll ein geglätteter Prozess diesen Mangel abstellen. Die bei der obigen Parameterschätzung für den geglätteten Fall ((Poly)Sar1archFit-Ausgabe) benutzten Werte  $\underline{w}_0 = -0.15$ ,  $\overline{w}_0 = 0.12$ ,  $\alpha = 0.5 = \gamma$  wurden aus dieser grafischen Analyse abgeleitet. Ganz analog zu der Situation im ungeglätteten Fall, werden wiederum 500 Realisationen des Smooth-AR(1)-ARCH(1)-Falls anhand ihrer empirischen Dichten mit den Allianz-Daten verglichen - siehe Abbildung 3.15. Im direkten Vergleich liegen die Kurven im Zentrum etwas niedriger, aber an den Rändern des Wertebereichs, wie erhofft, etwas höher. Dieser Eindruck wird bei einer angepassten Skala noch untermauert.

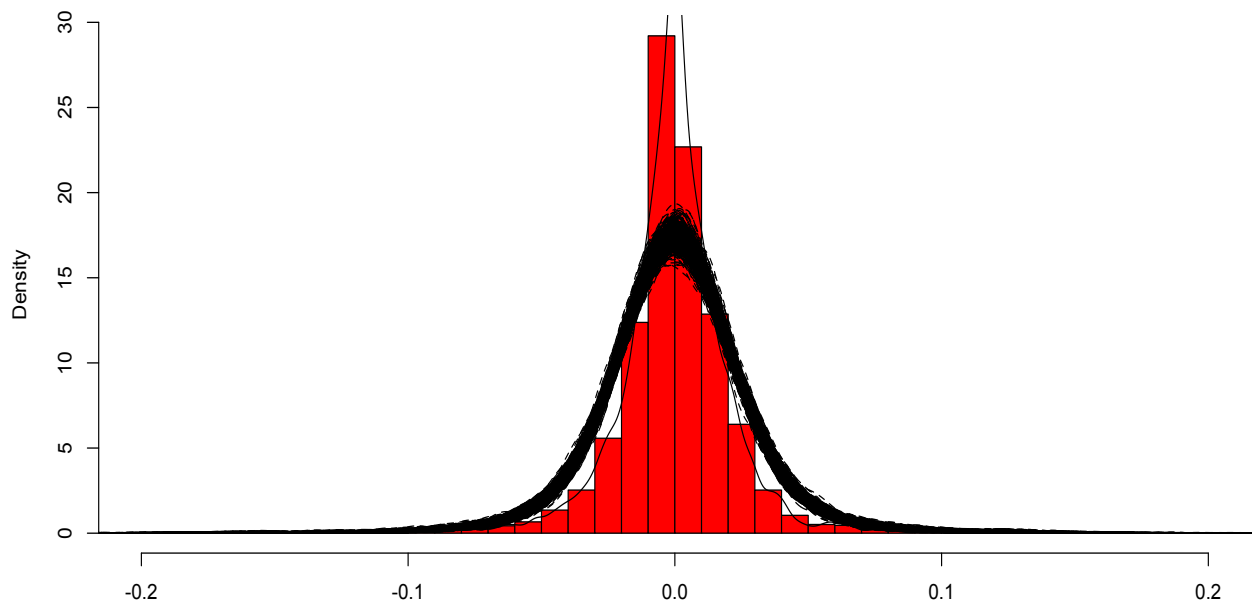


Abbildung 3.15: 500 geschätzte Dichten, zu 500 Realisationen eines Smooth-AR(1)-ARCH(1) Prozesses.

In der grafischen Umsetzung ist der gewünschte Effekt ganz klar zu erkennen. Ob sich damit, wie erhofft, die Anpassungsgüte verbessert hat, kann allerdings nicht anhand einer Grafik entschieden werden. In diesem Zusammenhang wird das Akaike Information Criterion (AIC) oder dessen Abwandlung Bayes Information Criterion (BIC) genutzt und baut auf dem bei der Parameterschätzung berechneten Log-Likelihood-Wert auf.



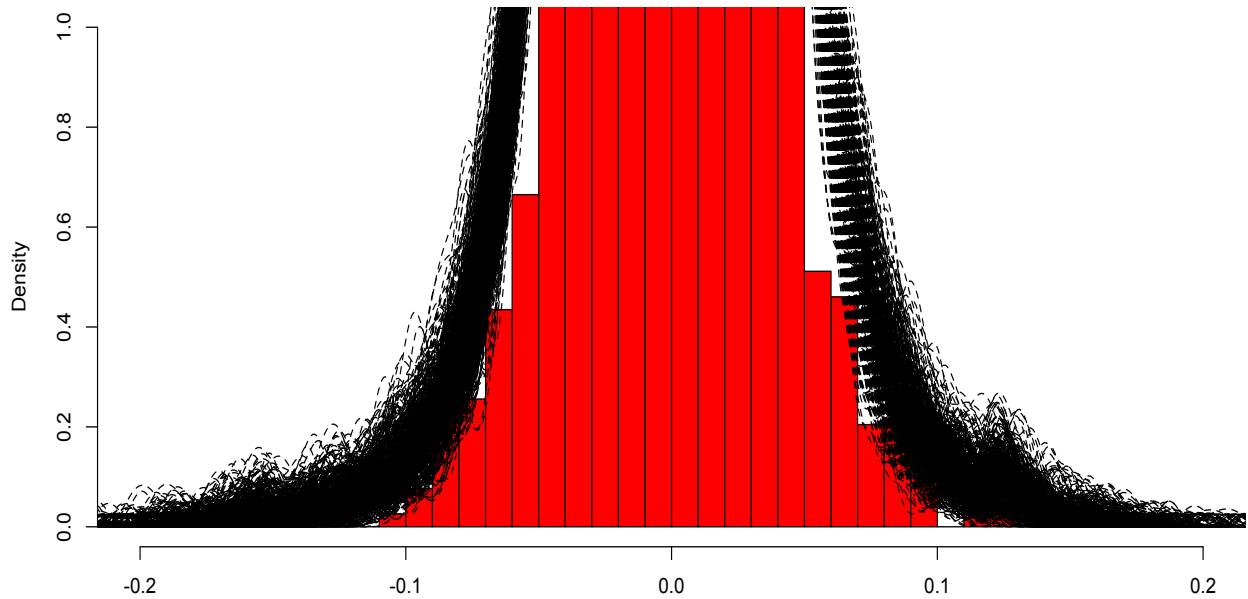


Abbildung 3.16: Ausschnitt zu Abbildung 3.15

### Definition 3.2. (*AIC*, *BIC*)

*Das Akaike Information Criterion*

$$AIC = 2K - 2\ln(L)$$

*gibt in einer Familie von Modellen für einen Datensatz, demjenigen mit minimalem AIC-Wert den Vorrang. Hierbei ist  $K$  die Anzahl der Parameter und  $L$  die maximale Likelihoodfunktion für das Modell.*

*Das Bayes Information Criterion*

$$BIC = \ln(N) \cdot K - 2\ln(L)$$

*ist eine Abwandlung, die durch Einbeziehung der Stichprobengröße  $N$ , der Bevorteilung von vielen Parametern entgegenwirkt.*

Nutzt man die `garchFit` Prozedur zur Parameterschätzung

```
> fit<-garchFit(...)
```

so kann durch den Aufruf von

```
> summary(fit)
```

auch der Wert für AIC und BIC ausgerechnet werden. Im AR(1)-ARCH(1)-Fall liegen die Werte bei

$$LLH = 9829.627, \quad \text{normalized} = 2.513971, \quad AIC = -5.025896, \quad BIC = -5.019480.$$

Unsere `ar1arch1Fit`-Prozedur liefert uns erst einmal nur den LLH-Wert (Log-Likelihood), der im übrigen mit der `garchFit`-Prozedur übereinstimmt und daraus lassen sich die restlichen Werte wie folgt berechnen:

$$\text{AIC} = \frac{2 \cdot 4}{3910} - 2 \cdot \frac{\text{LLH}}{3910}, \quad \text{BIC} = \frac{\ln(3910) \cdot 4}{3910} - 2 \cdot \frac{\text{LLH}}{3910}$$

Wobei `normalized` =  $\frac{\text{LLH}}{3910}$  und 3910 = Länge des Datensatzes, der durch `diff(log(Allianz))` entsteht.

Mit dem AIC/BIC-Wert lassen sich nun verschiedene Modelle in Hinblick auf ihre Anpassungsgüte vergleichen und damit eine Rangfolge erstellen. Nicht zu vergessen ist aber der Umstand, dass AIC/BIC nur ein relatives und kein absolutes Ergebnis liefert. Man kann also keine Aussage darüber treffen, ob ein Modell gut angepasst ist an die Daten (dazu siehe Residuen, QQ-Plot oben), sondern nur darüber ob es besser angepasst ist als ein Vergleichsmodell.

Da die Berechnung nur auf dem LLH-Wert und der Parameterzahl basiert, kann man Sie im `Sar1arch1Fit`-Fall analog durchführen und kommt mit der Festlegung

$$\alpha = \gamma = 0.5, \quad \overline{w_0} = 0.12, \quad \underline{w_0} = -0.15$$

zu den Ergebnissen

LLH = 9830.3487, normalized = 2.514155678, AIC = -5.02626532, BIC = -5.019849675.

Gemäß der obigen Definition, ist der AIC/BIC-Wert in diesem Fall niedriger und dem geglätteten Modell der Vorrang zu geben oder anders ausgedrückt, das geglättete Modell hat eine bessere Anpassung an die Allianz-Daten als das ungeglättete. Der Unterschied ist nicht wirklich groß, jedoch geht es hier erst einmal um das prinzipielle Ergebnis und eine Rechtfertigung für den Glättungsansatz.

Im `PolySar1arch1Fit`-Fall ergibt sich

LLH = 9830.3357, normalized = 2.514152353, AIC = -5.02625867, BIC = -5.019843025

und somit ein Ergebnis zwischen den Vorangegangenen. Der Glättungsansatz mittels Geradenstücken wäre hier, wenn auch nur knapp, die bessere Wahl.

### 3.4.2 Predict

Nicht nur bei Finanzmarktdaten, sondern bei so gut wie jeder über die Zeit hinweg betrachteten Größe (Niederschlagsmengen, Verkaufszahlen,...), ist man daran interessiert, ausgehend von den erhobenen Daten etwas über die Zukunft zu sagen. Mit einer gelungenen Parameterschätzung hat man ein Modell an der Hand, mit dessen Hilfe man die zukünftigen Werte mit hinreichender Güte beschreiben kann. Unsere Vergleichsprozedur `garchFit` liefert, mit der `predict`<sup>15</sup>-Funktion, standardmässig<sup>16</sup> eine Vorhersage über die 10 nächsten Werte.

<sup>15</sup><https://svn.r-project.org/Rmetrics/trunk/fGarch/R/methods-predict.R>

<sup>16</sup>Für kürzere oder längere Prognosen, siehe `Help`-Funktion.

Die Befehlsfolge in R

```
>x<-diff(log(Allianz))
>ar1arch1<-garchFit(~arma(1,0)+garch(1,0),data=x)
>predict(ar1arch1)
```

liefert folgendes Ergebnis:

	meanForecast	standardDeviation
1	0.0004684543	0.01708573
2	-0.0001896179	0.01992236
3	-0.0002803935	0.02095729
4	-0.0002929152	0.02136021
5	-0.0002946425	0.02152045
6	-0.0002948808	0.02158470
7	-0.0002949136	0.02161054
8	-0.0002949182	0.02162094
9	-0.0002949188	0.02162513
10	-0.0002949189	0.02162682

Für eine Umsetzung auf die Smooth-AR(1)-ARCH(1)-Situation, bedarf es zunächst einer genauen Vorstellung über die Entstehung der obigen Zahlen und ihrer Aussagekraft. Unser Ausgangsdatenmaterial umfasst 3911 `Allianz`-Datenpunkte, die nach den oben beschriebenen und begründeten Transformationen die 3910 `x`-Datenpunkte für die Parameteranpassung bilden. Aus den Ergebnissen von `garchFit`, werden folgende Werte für die Berechnung benötigt bzw. von `predict` übernommen:

<code>mu</code>	<code>beta</code>	<code>alpha0</code>	<code>alpha1</code>	<code>x[3910]</code>
-0.0002542373	0.1379416640	0.0002792524	0.4030117587	0.00523911
<code>x[3909]</code>	<code>z[3910]</code>			
-0.000823638	0.005606961			

Gemäß der Definition für AR(1)-ARCH(1)

$$W_t = \mu + \beta W_{t-1} + \sqrt{\alpha_0 + \alpha_1 W_{t-1}^2} \varepsilon_t$$

mit  $E(\varepsilon_t) = 0$  und  $E(\varepsilon_t^2) = 1$ , ergibt sich der zu erwartende Wert (`meanForecast`)  $\hat{W}_t = E(W_t | W_{t-1} = w) = \mu + \beta w$ , da nur der AR(1)-Teil für den Erwartungswert verantwortlich zeichnet (vgl. Tsay [31] 2.4.4). Der Parameter  $\mu$  entstammt dabei der Parameterschätzung.

Im konkreten Fall bedeutet dies, für den ersten zu schätzenden Wert

$$x[3911] = \mu + \beta \cdot x[3910] = -0.0002542373 + 0.1379416640 \cdot 0.00523911 = 0.0004684543.$$

Durch sukzessives Einsetzen, ergeben sich die weiteren Werte

$$x[3912] = \mu + \beta \cdot x[3911] = -0.0002542373 + 0.1379416640 \cdot 0.0004684543 = -0.0001896179$$

$x[3913] = \dots$

Der `meanForecast` ist dabei nichts anderes als ein Punktschätzer, der den tatsächlichen Wert vermutlich nicht genau treffen wird, aber zumindest in dessen Umkreis angesiedelt ist.

Die `standardDeviation` (Standardabweichung) beschreibt einen zugehörigen Bereich, um diese Punktschätzung herum. Ganz anschaulich will man die zufallsabhängige Streuung, die den Werten inne wohnt, greifbar machen. Der erste Eintrag beruht hierbei auf den vergangenen Werten

$$0.01708573 = \sqrt{w + \alpha_1 \cdot z[3910]^2} = \sqrt{0.0002792524 + 0.4030117587 \cdot (0.005606961)^2}$$

mit

$$z[3910] = x[3910] - \mu - \beta \cdot x[3909]$$

und die restlichen Werte ergeben sich wiederum durch sukzessives Einsetzen:

$$0.01992236 = \sqrt{0.0002792524 + 0.4030117587 \cdot (0.01708573)^2}$$

Entscheidend ist hierbei, die Berechnungsvorschrift nachzuvollziehen. Die Ausgangssituation stellt sich so dar, dass bis zum Zeitpunkt  $t=3910$  Daten vorliegen, man kann also bis zu diesem Zeitpunkt die Residuen einer AR(1)-Anpassung berechnen (siehe  $z[3910]$ ). Diese sind zu interpretieren als die zufallsbedingte Störung oder wie die Differenz zwischen dem `meanForecast` (Schätzwert) und dem tatsächlich angenommenen Wert. Auf Grund dieser Abweichung, macht man sich eine Vorstellung von der Streuungsbreite im Zeitpunkt  $t=3910$ . Dies ist nur eine Schätzung, da man bei einer einzigen Realisation keine Messung der tatsächlichen Streuung vornehmen kann. Für alle zukünftigen Zeitpunkte  $t \geq 3911$ , fehlt diese Vergleichsmöglichkeit jedoch und man muss zu einer anderen Methode übergehen, um die zu erwartende Streuung zu beschreiben. Gemäß der naiven Vorstellung, dass man ein zeitlich weiter entferntes Ereignis mit weniger Sicherheit beschreiben kann, sollte der Fehler zum einen anwachsen und zum anderen auf dem letzten zu berechnenden Wert aufbauen.

Dieser Idee und der Definition des AR(1)-ARCH(1)-Prozesses folgend, zeichnet nur der ARCH(1)-Term für die Streuung verantwortlich (vgl. Tsay [31] 3.4.3) und man erhält die Schätzwerte durch die Vorhersage eines ARCH(1)-Prozesses.

Für die Umsetzung auf das (Poly)-Smooth-AR(1)-ARCH(1)-Beispiel, sollte man sich an die Konstruktion erinnern, die diesen Prozess auszeichnet. Zulässige Glättungsfunktionen  $g$  wurden oben behandelt und führten in Komposition mit der Übergangsfunktion eines AR(1)-ARCH(1)-Modells zu diesem neuen Prozesstyp.

Übergangsfunktion AR(1)-ARCH(1):

$$r(w, x) = \beta w + \sqrt{\alpha_0 + \alpha_1 w^2} x$$

Wobei  $0 < \beta < 1$ ,  $\alpha_0 > 0$ ,  $\alpha_1 \geq 0$  und  $x$  als Realisation eines i.i.d.  $N(0, 1)$ -Prozesses.

Komposition von  $g$  und  $r$

$$W \times X \xrightarrow{r} \mathbb{R} \xrightarrow{g} W = \mathbb{R}_{\geq 0}$$

$$u(w, x) = g(r(w, x)).$$

Hier steht man nun vor dem Dilemma, dass die Dichte sich konstruktionsgemäß in drei Teile aufspaltet:

**Smooth-AR(1)-ARCH(1) (vgl. oben):**

$$f_{\hat{g} \circ X}(y) = \begin{cases} f_X(\hat{g}_1^{-1}(y)) \cdot \left| \frac{\partial \hat{g}_1^{-1}(y)}{\partial y} \right| = \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((\frac{y - \underline{w}_0}{\alpha} + \underline{w}_0) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\alpha} \right| & \text{für } y < \underline{w}_0 \\ f_X(\hat{g}_2^{-1}(y)) \cdot \left| \frac{\partial \hat{g}_2^{-1}(y)}{\partial y} \right| = \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(y - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot |1| & \text{für } \underline{w}_0 < y < \overline{w}_0 \\ f_X(\hat{g}_3^{-1}(y)) \cdot \left| \frac{\partial \hat{g}_3^{-1}(y)}{\partial y} \right| = \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((\frac{y - \overline{w}_0}{\gamma} + \overline{w}_0) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\gamma} \right| & \text{für } \overline{w}_0 < y. \end{cases}$$

**Poly-Smooth-AR(1)-ARCH(1) (vgl. oben):**

$$f_{\bar{g} \circ X}(y) = \begin{cases} f_X(\bar{g}_1^{-1}(y)) \cdot \left| \frac{\partial \bar{g}_1^{-1}(y)}{\partial y} \right| = \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(w_0 + 1 - (w_0 + 1 - y)^{\frac{1}{\alpha}} - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\alpha} (w_0 + 1 - y)^{\frac{1 - \alpha}{\alpha}} \right| & \text{für } y < \underline{w}_0 \\ f_X(\bar{g}_2^{-1}(y)) \cdot \left| \frac{\partial \bar{g}_2^{-1}(y)}{\partial y} \right| = \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(y - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot |1| & \text{für } \underline{w}_0 < y < \overline{w}_0 \\ f_X(\bar{g}_3^{-1}(y)) \cdot \left| \frac{\partial \bar{g}_3^{-1}(y)}{\partial y} \right| = \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((y - \overline{w}_0 + 1)^{\frac{1}{\gamma}} + \overline{w}_0 - 1 - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\gamma} (y - \overline{w}_0 + 1)^{\frac{1 - \gamma}{\gamma}} \right| & \text{für } \overline{w}_0 < y. \end{cases}$$

Durch die Komposition mit  $g$ , beraubt man sich einiger wertvoller Eigenschaften des Ausgangsprozesses. So verschwindet die klare Trennung zwischen AR-Teil und ARCH-Teil. Man kann beispielsweise für eine Punktschätzung wie oben, nicht außer Acht lassen, dass ihr konkreter Wert sich in Abhängigkeit von dem getroffenen Gebiet (unterhalb, innerhalb oder oberhalb des Kompaktums) auf unterschiedliche Art und Weise errechnet. Dies führt unweigerlich dazu, dass man für die Vorhersage von zukünftigen Werten auf einen anderen Ansatz zurückgreifen muss.

Das prinzipielle Problem findet sich auch in wesentlich „leichteren“ Begleitumständen, wie etwa bei dem Threshold-AR-Modell und dessen Vorhersagemöglichkeit (`predict.TAR`). Die folgende Idee ist also angelehnt an die Vorgehensweise im TAR-Fall und die in R durch das Paket TSA<sup>17</sup> zur Verfügung gestellten Prozeduren. Da eine tiefere Auseinandersetzung mit den konkreten Algorithmen nur vom eigentlichen Ziel ablenken würde, sei es dem interessierten Leser überlassen, sich die Funktionen `tar`, `predict.TAR` und `tar.sim` mit der Hilfefunktion<sup>18</sup> von R genauer anzusehen.

Wir wollen uns an dieser Stelle damit begnügen, die Vorgehensweise auf das (Poly)-Smooth-AR(1)-ARCH(1)-Beispiel umzusetzen. Dies ist auch wiederum ein Punkt, an dem man eine Anknüpfung zu den vorangegangenen Kapiteln findet. So bedient man sich der Parameterschätzungs- und Simulations-Algorithmen, um darauf aufbauend weiterführende Prozeduren zu konstruieren. Dadurch zeigt sich erneut und sehr eindrucksvoll die Notwendigkeit des Vorangegangenen.

Um eine Vorhersage zu treffen, muss man den vorgeschlagenen Wert auch mit einem gewissen „Gewicht“ ausstatten oder vielleicht besser gesagt, mit einer Rechtfertigung für seine ihm zugesprochene Rolle.

Die Vorgehensweise der `Predict-(Poly)-Smooth-ar1arch1`-Prozedur lässt sich wie folgt umschreiben:

Zu gegebenem Datenmaterial erhält man aus der Parameterschätzung die notwendigen Werte, um es durch eine Simulation in die Zukunft fortschreiben zu können. Den zukünftigen Werten liegt damit natürlich eine Gesetzmässigkeit zu Grunde, jedoch sind sie zum Teil zufallsabhängig und damit einer gewissen Streuung unterworfen. Eben dieser Umstand führte im ungeglätteten Fall dazu, dass man den am wahrscheinlichsten zu erwartenden Wert als Vorhersage gewählt hat. Die Möglichkeit zu dieser Wahl entsprang dabei, wie gesehen, der Konstruktion des AR(1)-ARCH(1)-Modells.

Das Glätten verhindert jedoch einen so schnellen Einblick und macht deshalb einen kleinen Umweg nötig. Wenn sich der Wert also nicht direkt berechnen lässt oder nur unter großem Aufwand, kann man die Simulation dazu nutzen, um eine angemessene Anzahl von möglichen Kandidaten für den nächsten Datenpunkt zu generieren. Aus dieser neuen Datenmenge leitet man nun die Vorhersage ab, indem man den Median darüber bildet. Diese Wahl wird umso sinnvoller, je mehr Kandidaten man mit einbezieht.

Aus der Parameterschätzung und Simulation kann man auf diese Art und Weise eine `Predict`-Funktion konstruieren. Für den Smooth-AR(1)-ARCH(1)-Fall ergibt sich die Einschrittvorhersage, gemäß der vorangegangenen Ergebnisse durch folgendes Minimalprogramm:

---

<sup>17</sup><http://cran.r-project.org/web/packages/TSA/TSA.pdf>

<sup>18</sup>z.B.: `?tar.sim` oder `help("tar.sim")`

## Predict-Smooth-ar1arch1-1Step

```
set.seed(1)
eps<-rnorm(1000)
mu<- -0.000257695
beta<-0.138581
alpha0<-0.000279295
alpha1<-0.404023
alpha<-0.5
gamma<-0.5
w0unten<- -0.15
w0oben<-0.12
y=rep(0,1000)
for(i in 1:1000){
y[i]=mu+beta*x[length(x)]+sqrt(alpha0+alpha1*x[length(x)]^2)*eps[i]
if(y[i]>w0oben)y[i]<-w0oben+gamma*(y[i]-w0oben)
if(y[i]<w0unten)y[i]<-w0unten+alpha*(y[i]-w0unten)
y[i]<-y[i]}
z=median(y)
```

Hierbei werden auf Grundlage der Parameterschätzung (`mu`, `beta`, `alpha0`, `alpha1`) und des letzten Datenpunktes (`x[length(x)]`), genau 1000 mögliche Kandidaten für den nächsten Wert erzeugt und anschliessend der Median gebildet.

Als Vorhersage für den nächsten Datenpunkt erhält man: `-0.0001336027`

Eine Verallgemeinerung des obigen Algorithmus findet sich im Anhang (Abschnitt - 4.9). Die Möglichkeit verschiedene Glättungsfunktionen bei unterschiedlichen Glättungsparametern anzusetzen, sowie den Vorhersagehorizont zu erhöhen, bietet einige Erweiterungsmöglichkeiten. Diese Variabilität führt jedoch unweigerlich zu einem unübersichtlicheren und verschachtelten Algorithmus, der zum Verständnis der Grundidee an dieser Stelle keineswegs nötig ist.

Zusammenfassend hat man mit der `Predict`-Prozedur nicht nur den Abschluss der empirischen Untersuchung im Glättungsfall vor sich, sondern eben auch die Zusammenführung der vorangegangenen Ergebnisse. Vieles wurde zwar exemplarisch am AR(1)-ARCH(1)-Modell erarbeitet, besitzt in der Regel jedoch allgemeine Gültigkeit. Der runde Abschluss der Untersuchungen in der vorliegenden Arbeit, sollte zugleich auch der Startpunkt für analoge Umsetzungen auf andere autoregressive Modelle unter Glättungsaspekten sein. Eine endgültige Beurteilung des Glättungsansatzes wird sich nicht hier und jetzt bewerkstelligen lassen, sondern erst durch die konsequente Umsetzung in verschiedenen Modellkonstellationen erfolgen können.

## 4 Anhang

Einige der nun folgenden Resultate sind von zentraler Bedeutung für das Vorangegangene bzw. dessen Verständnis, wie die mehrfachen Verweise auf den Anhang deutlich machen. Für den besseren Lesefluss und eine überschaubare Darstellung der Kernaussagen, war es jedoch unausweichlich einiges gesondert zusammenzustellen. Prinzipiell würde es sich natürlich anbieten, den jeweiligen Abschnitt genau dann zu lesen, wenn er als Quelle genannt wird. Die Unterbrechung des Leseflusses und auch die Mehrfacherwähnung machen es jedoch notwendig, den Anhang auch für sich „lesbar“ zu machen. Zwar sind die einzelnen Teilabschnitte nicht immer in einen sinnvollen Zusammenhang oder Reihenfolge zu bringen, bieten jeder für sich aber einen tieferen Einblick in das Behandelte.

### 4.1 Kontraktionsbedingungen in 1 und 2 Schritt Übergangsfunktionen

Die Kontraktionsbedingungen wurden ganz zu Beginn als zentrales Kriterium für IFS herausgestellt und eine mögliche Abschwächung durch eine Mehrfachiteration mittels der Submultiplikativität motiviert. Um jegliche Zweifel aus dem Weg zu räumen und auch Hinweise in der Arbeit von Straumann [28] zu untermauern, sollen konkrete Beispiele herangezogen werden. Bei verschiedenen Autoren, vgl. etwa Iosifescu [16] oder Straumann [28], wird die Kontraktivität erst nach  $n$  Übergängen untersucht, d.h.

$$u^n(w, x_1, \dots, x_n) := u(\dots u(u(w, x_1), x_2), \dots, x_n).$$

Hier soll an zwei Beispielen verdeutlicht werden, dass es tatsächlich nicht immer ausreicht nur den 1 Schritt Übergang zu betrachten.

#### 4.1.1 IFS

In diesem Abschnitt wird ein IFS konstruiert, welches die 1 Schritt Kontraktionsbedingung verfehlt, aber die 2 Schritt Bedingung erfüllt. Mit anderen Worten

$$\sup_{\substack{w' \neq w'' \\ w', w'' \in W}} \int_X \frac{d(u(w', x), u(w'', x))}{d(w', w'')} p(dx) > 1,$$

aber

$$\sup_{\substack{w' \neq w'' \\ w', w'' \in W}} \int_{X^2} \frac{d(u^2(w', x), u^2(w'', x))}{d(w', w'')} p_2(dx^{(2)}) < 1,$$

wobei  $p_2$  die Randverteilung von  $(X_t, X_{t+1})$  für  $t \in \mathbb{Z}$  bzw.  $t \in \mathbb{N}$  (vgl. [16] - (3.6), (3.15),  $n=2$ ,  $\alpha = 1$ ).



**Beispiel 1:**

Man betrachte das IFS mit  $W = [0, 1]$ ,  $X = \{1, 2\}$ ,  $p = (p_i)_{1 \leq i \leq 2}$  ( $p_1 = 2/3, p_2 = 1/3$ ) und  $u(w, x)$  als  $u_i(w), i = 1, 2$ :

$$\begin{aligned} u_1(w) &:= \frac{2}{3}w^2 \\ u_2(w) &:= \frac{3}{5} + \frac{2}{5}w \end{aligned}$$

Diese Festlegung der Selbst-Abbildungen  $u_i : W \rightarrow W$ ,  $1 \leq i \leq 2$  und  $d = |\cdot|$  als Betragsfunktion auf  $W$ , führen zu:

$$\begin{aligned} & \sup_{\substack{w' \neq w'' \\ w', w'' \in W}} \int_X \frac{d(u_x(w'), u_x(w''))}{d(w', w'')} p(dx) \\ &= \sup_{\substack{w' \neq w'' \\ w', w'' \in W}} \sum_{i=1}^2 p_i \frac{|u_i(w') - u_i(w'')|}{|w' - w''|} \\ &= \sup_{\substack{w' \neq w'' \\ w', w'' \in W}} \left\{ \frac{2}{3} \frac{|\frac{2}{3}w'^2 - \frac{2}{3}w''^2|}{|w' - w''|} + \frac{1}{3} \frac{|\frac{3}{5} + \frac{2}{5}w' - (\frac{3}{5} + \frac{2}{5}w'')|}{|w' - w''|} \right\} \\ &= \sup_{\substack{w' \neq w'' \\ w', w'' \in W}} \left\{ \frac{4}{9}|w' + w''| + \frac{2}{15} \right\} \\ &= \frac{92}{90} > 1 \end{aligned}$$

Für die 2 Schritt Situation müssen die Fälle  $(u_1 \circ u_1)$ ,  $(u_1 \circ u_2)$ ,  $(u_2 \circ u_1)$  und  $(u_2 \circ u_2)$  behandelt werden. Dies führt zu

$$\begin{aligned} & \sup_{\substack{w' \neq w'' \\ w', w'' \in W}} \left\{ \frac{1}{3} \frac{2}{3} \frac{|\frac{2}{5}(\frac{2}{3}w'^2) + \frac{3}{5} - (\frac{2}{5}(\frac{2}{3}w''^2) + \frac{3}{5})|}{|w' - w''|} + \frac{2}{3} \frac{1}{3} \frac{|\frac{2}{3}(\frac{3}{5} + \frac{2}{5}w')^2 - \frac{2}{3}(\frac{3}{5} + \frac{2}{5}w'')^2|}{|w' - w''|} \right. \\ & \quad \left. + \frac{2}{3} \frac{2}{3} \frac{|\frac{2}{3}(\frac{2}{3}w'^2)^2 - \frac{2}{3}(\frac{2}{3}w''^2)^2|}{|w' - w''|} + \frac{1}{3} \frac{1}{3} \frac{|\frac{3}{5} + \frac{2}{5}(\frac{3}{5} + \frac{2}{5}w') - (\frac{3}{5} + \frac{2}{5}(\frac{3}{5} + \frac{2}{5}w'') + \frac{3}{5})|}{|w' - w''|} \right\} \\ &= \frac{4748}{6075} < 1 \end{aligned}$$

Dieses kurze Beispiel verdeutlicht sehr eindrucksvoll, dass die Kontraktionsbedingung an  $u^2$  (bzw.  $u^n$ ) eine wirklich schwächere Bedingung darstellen kann. Für die SRE-Situation wird jetzt ein ähnliches Beispiel konstruiert. Die zentralen Unterschiede sind die veränderte Kontraktionsbedingung und die Abhängigkeit der betrachteten Abbildungen von  $t$ , d.h. ein Prozess  $(u_t)$ .

### 4.1.2 SRE

Im folgenden wird

$$\mathbb{E}[\log \Lambda(u_0)] < 0$$

verwendet, um die Existenz und Eindeutigkeit einer stationären Lösung von

$$Y_{t+1} = u_t(Y_t), \quad t \in \mathbb{Z}$$

sicherzustellen.

Das Hauptwerkzeug für den Beweis ist die „backward iteration“ oder Rückwärtsiteration. Hierbei wird die Beobachtung zu einem Zeitpunkt  $t$  verstanden, als Resultat einer in der Vergangenheit gestarteten Iteration.

Wiederum mit Submultiplikativität von  $\Lambda$  erahnt man die Abschwächung der Hauptbedingung zu

$$\mathbb{E}[\log \Lambda(u_0 \circ \dots \circ u_{-r+1})] < 0.$$

Es stellt sich die Frage nach einem Beispiel für die Situation  $\mathbb{E}[\log \Lambda(u_0)] \geq 0$ , aber trotzdem  $\mathbb{E}[\log(u_0 \circ \dots \circ u_{-r+1})] < 0$  für  $r \geq 2$ . Einen Hinweis darauf liefert Straumann [28] : „This is e.g. the case for the SRE arising from the state space representation of GARCH(p,q), see Section 3.3.1“, gibt jedoch kein Beispiel an.

Das AGARCH(p,q)-Modell (asymmetric generalized autoregressive conditionally heteroscedastic) umfasst als Spezialfall auch GARCH(p,q):

([28] S. 50 - Section 3.3.1)

$$\begin{aligned} X_t &= \sigma_t Z_t, \quad t \in \mathbb{Z} \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2, \quad t \in \mathbb{Z}, \end{aligned}$$

wobei die Parameter  $\alpha_i, \beta_i$  nichtnegativ und  $(Z_t)$  i.i.d. mit  $\mathbb{E}Z_0 = 0, \mathbb{E}Z_0^2 = 1$ .

Für  $p, q \geq 2$  ergeben die  $(p+q+1)$ -dimensionalen Vektoren

$$Y_t = (\sigma_t^2, \dots, \sigma_{t-q+1}^2, X_{t-1}^2, \dots, X_{t-p+1}^2)^T, \quad B = (\alpha_0, 0, \dots, 0)^T$$

und die  $(p+q+1) \times (p+q+1)$  Matrizen

$$A_t = \begin{pmatrix} \alpha_1 Z_t^2 + \beta_1 & \beta_2 & \dots & \beta_{q-1} & \beta_q & \alpha_2 & \alpha_3 & \dots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ Z_t^2 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}, \quad t \in \mathbb{Z}$$

für diese Modelle eine neue Darstellungsform

$$\begin{aligned} Y_{t+1} &= \psi_t(Y_t) \\ &= A_t Y_t + B \quad t \in \mathbb{Z} \end{aligned}$$

mit  $(A_t, B)$  i.i.d..

Es werden zwei unterschiedliche Normen herangezogen, um dieses Beispiel weiter zu behandeln:

Für eine Matrix  $A = (a_{ij}) \in \mathbb{R}^{d \times d}$  definiert man die Frobenius-Norm durch

$$\|A\|_F = \left( \sum_{i,j=1}^d a_{ij}^2 \right)^{1/2}.$$

Die Operatornorm (zugeordnet zur euklidischen Norm)

$$\begin{aligned} \|A\|_{op} &= \sup_{x \neq 0} \frac{|Ax|}{|x|} \\ &= \sqrt{\lambda_{\max}(A^H A)} \end{aligned}$$

( $A^H$  ist die konjugiert transponierte Matrix) wird auch Spektralnorm genannt.

Zusammen mit der Definition für  $\Lambda$  erkennt man

$$\Lambda(u_0 \circ \dots \circ u_{-r+1}) = \|A_0 \dots A_{-r+1}\|_{op}.$$

Der spezielle Aufbau von  $A_t$  und die hohe Anzahl von Einträgen gleich 1, geben eventuell die Möglichkeit die Beziehung von Frobenius und Operatornorm weiter auszunutzen

$$\frac{1}{\sqrt{d}} \|A\|_F \leq \|A\|_{op} \leq \|A\|_F \quad A \in \mathbb{R}^{d \times d}$$

$$\mathbb{E}[\log \frac{1}{\sqrt{d}} \|A\|_F] \leq \mathbb{E}[\log \|A\|_{op}] \leq \mathbb{E}[\log \|A\|_F]$$

und so etwas in Hinblick auf:  $\mathbb{E}[\log \|A_0\|_{op}] \geq 0$  aber  $\mathbb{E}[\log \|A_0 \circ \dots \circ A_{-r+1}\|_{op}] < 0$  abzuleiten.

### Beispiel 2:

Mit  $p = q = 2$  folgt

$$Y_t = (\sigma_t^2, \sigma_{t-1}^2, X_{t-1}^2)^t, \quad B = (\alpha_0, 0, 0)^t, \quad A_t = \begin{pmatrix} \alpha_1 Z_t^2 + \beta_1 & \beta_2 & \alpha_2 \\ 1 & 0 & 0 \\ Z_t^2 & 0 & 0 \end{pmatrix}$$

und für  $\alpha_0 = \frac{1}{10}$ ,  $\alpha_1 = \frac{1}{20}$ ,  $\alpha_2 = \frac{1}{30}$ ,  $\beta_1 = \frac{1}{5}$ ,  $\beta_2 = \frac{1}{15}$  ergibt sich

$$\|A_0\|_F = \left( \sqrt{\frac{1}{400} Z_0^2 + \frac{1}{50} Z_0^2 + \frac{1}{25} + \frac{1}{225} + \frac{1}{900} + 1 + Z_0^2} \right)$$

Da  $\sqrt{\cdot} \geq 1$ , kann man sich so eine obere Schranke konstruieren. Seien die  $Z_t$  i.i.d. und gleich verteilt auf  $\{-1, 1\}$ , so folgt:

$$\begin{aligned}\mathbb{E}[\log \|A_0\|_F] &= \frac{1}{2} \log \left( \sqrt{\frac{1}{400} - 1^4 + \frac{1}{50} - 1^2 + \frac{1}{25} + \frac{1}{225} + \frac{1}{900} + 1 - 1^4} \right) \\ &\quad + \frac{1}{2} \log \left( \sqrt{\frac{1}{400} 1^4 + \frac{1}{50} 1^2 + \frac{1}{25} + \frac{1}{225} + \frac{1}{900} + 1 + 1^4} \right) \\ &\approx 0.363304 > 0.\end{aligned}$$

Für dieses spezielle  $Z_t$  hängt die Matrix  $A_t$  gar nicht mehr von  $t$  ab und es folgt

$$\begin{aligned}\mathbb{E}[\log \|A_0\|_{op}] &= \mathbb{E}[\log \left\| \begin{pmatrix} \frac{1}{20} + \frac{1}{5} & \frac{1}{15} & \frac{1}{30} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \right\|_{op}] \\ &\approx 0.362\end{aligned}$$

aber

$$\mathbb{E}[\log \frac{1}{\sqrt{3}} \|A_0\|_F] \approx -0.186002.$$

In diesem Fall ergibt sich also nur die Abschätzung

$$- \leq + \leq +$$

und die Beziehung zwischen den Normen gibt keinen tieferen Einblick. Betrachtet man jedoch den Fall  $A_0 A_{-1} = A_0^2$

$$\begin{aligned}\mathbb{E}[\log \|A_0 A_{-1}\|_{op}] &= \mathbb{E}[\log \|A_0^2\|_{op}] \\ &= \mathbb{E}[\log \left\| \begin{pmatrix} \frac{13}{80} & \frac{1}{60} & \frac{1}{120} \\ \frac{1}{4} & \frac{1}{15} & \frac{1}{30} \\ \frac{1}{4} & \frac{1}{15} & \frac{1}{30} \end{pmatrix} \right\|_{op}] \\ &\approx -0.909526\end{aligned}$$

so erhält man das von Straumann [28] erwähnte Ergebnis. Die Situation ändert sich auch nicht mit höherer Dimension oder einer anderen Verteilung. Die Festlegung auf beispielsweise  $Z_t \sim \mathcal{N}(0, 1)$ ,  $\alpha_1 = \frac{1}{4}$ ,  $\alpha_2 = \frac{1}{6}$ ,  $\beta_1 = \frac{1}{5}$ ,  $\beta_2 = \frac{1}{5}$  ergibt ein vergleichbares Ergebnis. Im Allgemeinen kann man jedoch nicht wie erhofft die Frobeniusnorm als Indikator für die gewünschte Situation nutzen. Nichts desto trotz hat man mit obigen Beispielen eine greifbare Rechtfertigung der Abschwächung der Kontraktionsbedingung, die zuvor nur intuitiv über die Submultiplikativität begründet war.

## 4.2 Driftbedingung für den geglätteten Prozess

In Abschnitt 2.2 wurde der Effekt einer Glättungsfunktion auf die Driftbedingung für geometrische Ergodizität untersucht. Um die Gültigkeit der Driftbedingung auch im geglätteten Fall zu gewährleisten (vgl. Satz 2.7), wurde die Berechnung des Integrals für den Erwartungswert aufgespalten. Durch die Aufspaltung in die drei Teilbereiche (unterhalb, innerhalb und oberhalb des Kompaktums) wurden die beiden Integrale (ungeglättet, geglättet) miteinander vergleichbar, ohne jedoch den konkreten Wert berechnen zu müssen. Während sich die Aufspaltung im ungeglätteten Fall relativ schnell ergibt, bedarf es für die selbe Vorgehensweise im geglätteten Fall einer Anwendung des Transformationssatzes (Satz 2.8) und mehrerer kleinerer Anpassungen, wie etwa Substitutionen. Dieses längere Unterfangen hätte im Beweis zu Satz 2.7 nur die eigentliche Kernaussage aus dem Fokus gerückt und wurde deshalb in den Anhang verschoben.

Wenn man nun, wie unter (6) im Beweis zu Satz 2.7, zunächst die Betragsfunktion auf das Integral für den geglätteten Fall umsetzt, erhält man:

$$\begin{aligned}
 (7) = & \int_0^{\frac{w_0}{\alpha}} y^n \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} \left( e^{-\frac{((\frac{y-w_0}{\alpha} + w_0) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\alpha} \right| + e^{-\frac{(-(\frac{y-w_0}{\alpha} + w_0) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\alpha} \right| \right) dy \\
 & + \int_{\frac{w_0}{\alpha}}^{\frac{\overline{w_0}}{\alpha}} y^n \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} \left( e^{-\frac{(y - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} + e^{-\frac{(-(\frac{y-w_0}{\alpha} + w_0) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\alpha} \right| \right) dy \\
 & + \int_{\frac{\overline{w_0}}{\alpha}}^{\infty} y^n \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} \left( e^{-\frac{((\frac{y-\overline{w_0}}{\gamma} + \overline{w_0}) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\gamma} \right| + e^{-\frac{(-(\frac{y-w_0}{\alpha} + w_0) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\alpha} \right| \right) dy
 \end{aligned}$$

Die Aufspaltung der Integrationsbereiche ist hier bereits gegeben. Stellt man jedes der drei Teilintegrale wiederum als Summe zweier Integrale dar und sortiert sie in die „richtige“ Reihenfolge, so erhält man das geglättete Gegenstück zu dem für (6) vielleicht etwas merkwürdig erscheinenden Konstrukt:

$$\begin{aligned}
(7) &= \int_0^{\underline{w_0}} y^n \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((\frac{y-\underline{w_0}}{\alpha} + \underline{w_0}) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\alpha} \right| dy & (i^*) \\
&+ \int_{\underline{w_0}}^{\overline{w_0}} y^n \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(y - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dy & (ii^*) \\
&+ \int_{\overline{w_0}}^{\infty} y^n \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((\frac{y-\overline{w_0}}{\gamma} + \overline{w_0}) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\gamma} \right| dy & (iii^*) \\
&+ \int_0^{\underline{w_0}} y^n \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(-(\frac{y-\underline{w_0}}{\alpha} + \underline{w_0}) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\alpha} \right| dy & (iv^*) \\
&+ \int_{\underline{w_0}}^{\overline{w_0}} y^n \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(-(\frac{y-\underline{w_0}}{\alpha} + \underline{w_0}) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\alpha} \right| dy & (v^*) \\
&+ \int_{\overline{w_0}}^{\infty} y^n \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(-(\frac{y-\overline{w_0}}{\gamma} + \overline{w_0}) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} \cdot \left| \frac{1}{\alpha} \right| dy & (vi^*)
\end{aligned}$$

Damit ähneln sich (6) und (7) schon sehr stark, aber ein direkter Vergleich scheitert immer noch an den unterschiedlichen Termen in der Exponentialfunktion. Um diesem Problem habhaft zu werden, substituiert man in den Fällen (i\*), (iv\*), (v\*), (vi\*) mit

$$\frac{y - \underline{w_0}}{\alpha} + \underline{w_0} = z, \quad \frac{1}{\alpha} dy = dz,$$

sowie im Fall (iii\*) mit

$$\frac{y - \overline{w_0}}{\gamma} + \overline{w_0} = t, \quad \frac{1}{\gamma} dy = dt.$$

Da ja  $\left| \frac{1}{\alpha} \right| = \frac{1}{\alpha}$  und  $\left| \frac{1}{\gamma} \right| = \frac{1}{\gamma}$ , erhält man nach der Substitution wiederum eine leicht veränderte Darstellung:

$$\begin{aligned}
(7) &= \int_{\frac{-w_0}{\alpha} + \underline{w_0}}^{\underline{w_0}} (\alpha \cdot (z - \underline{w_0}) + \underline{w_0})^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(z - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dz & (\text{i}^{**}) \\
&+ \int_{\underline{w_0}}^{\underline{w_0}} y^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(y - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dy & (\text{ii}^{**}) \\
&+ \int_{\underline{w_0}}^{+\infty} (\gamma \cdot (t - \underline{w_0}) + \underline{w_0})^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(t - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dt & (\text{iii}^{**}) \\
&+ \int_{\frac{-w_0}{\alpha} + \underline{w_0}}^{\underline{w_0}} (\alpha \cdot (z - \underline{w_0}) + \underline{w_0})^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((-z) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dz & (\text{iv}^{**}) \\
&+ \int_{\frac{\underline{w_0} - \underline{w_0}}{\alpha} + \underline{w_0}}^{\underline{w_0}} (\alpha \cdot (z - \underline{w_0}) + \underline{w_0})^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((-z) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dz & (\text{v}^{**}) \\
&+ \int_{\frac{\underline{w_0} - \underline{w_0}}{\alpha} + \underline{w_0}}^{+\infty} (\alpha \cdot (z - \underline{w_0}) + \underline{w_0})^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((-z) - \beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dz & (\text{vi}^{**})
\end{aligned}$$

In der Form  $\{(i), \dots, (vi)\}$  und  $\{(i^{**}), \dots, (vi^{**})\}$ , lassen sich (6) und (7) schon wesentlich besser miteinander vergleichen, während dies für  $\{(i), \dots, (vi)\}$  und  $\{(i^*), \dots, (vi^*)\}$  eben nicht möglich ist. Über diesen Umweg werden dann letztendlich doch Schlussfolgerungen zum Verhältnis von  $\{(i), \dots, (vi)\}$  und  $\{(i^*), \dots, (vi^*)\}$  gemacht bzw. über den bedingten Erwartungswert im ungeglätteten und geglätteten Fall. Bemerkenswert an dieser Stelle ist vielleicht auch der Umstand, dass bei  $\alpha \rightarrow 1$  und  $\gamma \rightarrow 1$ , sich  $\{(i^{**}), \dots, (vi^{**})\}$  und  $\{(i), \dots, (vi)\}$  bzw.  $\{(i^*), \dots, (vi^*)\}$  und  $\{(i), \dots, (vi)\}$  immer weiter annähern. Man kann buchstäblich die Auswirkungen einer „schwächer“ werdenden Glättungsfunktion an dieser Darstellung ablesen.

Die Dichtefunktion beschreibt ganz anschaulich, wie die Wahrscheinlichkeitsmasse sich über den Definitionsbereich verteilt. Mit dieser Vorstellung kann man sich die Beweisidee deutlich machen und auch die Grafiken zur Plausibilitätsüberlegung genauer betrachten (Abschnitt 4.3), denn es liegt in beiden Fällen der gleiche Gedanke zu Grunde. Durch die Betragsfunktion wird die gesamte Wahrscheinlichkeitsmasse, die zuvor auf  $\mathbb{R}_-$  entfallen war, auf  $\mathbb{R}_+$  umgelegt. Die Terme (iv), (v) und (vi) bzw. ihre Entsprechungen mit (\*) oder (\*\*) beschreiben eben genau diese umgelegte Wahrscheinlichkeitsmasse. Wenn man den Erwartungswert von  $|X|$  berechnet, mit Hilfe der Dichte  $f_X(x)$ , so ergibt sich das selbe Ergebnis wenn man den Erwartungswert von  $\tilde{X}$  berechnet, wobei die zu  $\tilde{X}$  gehörige Dichte eben durch die Übertragung der Wahrscheinlichkeitsmasse hervorgegangen ist. Man macht also nichts anderes, als die Betragsfunktion bereits auf die Dichte wirken zu lassen und muss dann bei der Berechnung des Erwartungswertes nicht mehr ein Integral von  $-\infty$  bis  $+\infty$  betrachten, sondern nur noch von 0 bis  $+\infty$ . Ganz analog verhält es sich dann mit  $X^\eta$  bzw.  $\tilde{X}^\eta$ . In (6) und (7) haben wir also nur eine Darstellung für

$\mathbb{E}|X|^\eta = \sum(i)$  bzw.  $\mathbb{E}|\hat{g} \circ X|^\eta = \sum(i^{**})$  zu gewinnen versucht. Im Rahmen der Verschiebung der Wahrscheinlichkeitsmasse wird jetzt geschlussfolgert, dass die Glättungsfunktion  $\hat{g}$  bereits vor dem Anwenden der Betragsfunktion zu einer Verschiebung der Wahrscheinlichkeitsmasse führt und eben diese gilt es zu interpretieren, um Aussagen über den geglätteten Fall machen zu können. Für die Glättungsfunktion  $g$  bzw.  $g^*$ , hätte die Betragsfunktion sogar keine Wirkung, da diese Glättungsfunktionen ohnehin nur  $\mathbb{R}_+$  als Bildbereich haben.

### Schlussfolgerung:

Schaut man über die unterschiedlichen Variablen  $x, y, t, z$  hinweg, so erkennt man in den 6 Teilintegralen jeweils die selbe Dichte (d.h. z.B. in (i) und (i<sup>\*\*</sup>) taucht beide Male der Term

$$\frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(x-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}}$$

auf). Die Integrale unterscheiden sich demnach nur in dem Vorfaktor, mit dem die Dichte jeweils multipliziert wird bzw. im Integrationsbereich. In einem letzten Schritt, wird nun noch der Integrationsbereich angeglichen, so dass für eine Interpretation nur noch der Vorfaktor beachtet werden muss. Es reicht hierbei (i<sup>\*\*</sup>) und (iv<sup>\*\*</sup>) anzupassen, da die restlichen Integrale in diesem Punkt bereits übereinstimmen bzw. durch Zusammenfassen die Unterschiede im Integrationsbereich wegfallen ((v<sup>\*\*</sup>) + (vi<sup>\*\*</sup>) und (v) + (vi)). Hierbei werden zum Zwecke der Übersichtlichkeit auch die Variablen angeglichen.

Die Idee dahinter ist ganz einfach die, dass man über dem jeweils identischen Integrationsbereich die identische Dichte betrachtet und ganz allein der Vorfaktor als eine Art Positionskoordinate verstanden werden kann. Im ungeglätteten Fall wird ja für die Erwartungswertbildung die Dichte  $f(x)$  mit  $x$  multipliziert. Im geglätteten Fall wird man nach der Anpassung des Integrationsbereichs zwar die selbe Dichte über den jeweiligen Teilintervallen betrachten, diese aber mit einem anderen Vorfaktor multiplizieren. Ein größerer Vorfaktor bedeutet, dass die Dichte nach rechts (entlang der  $x$ -Achse) verschoben wird und ein kleinerer Vorfaktor bedeutet demnach eine Verschiebung nach links.

Eine derartige Verschiebung kann in Hinblick auf die Erwartungswertberechnung dahingehend interpretiert werden, dass bei einer Verschiebung der Dichte die Wahrscheinlichkeitsmasse in diese Richtung bewegt wird und damit auch der Erwartungswert sich in die gleiche Richtung bewegen muss.



$$\begin{aligned}
(7) &= \int_0^{\underline{w}_0} (\alpha \cdot (x - \underline{w}_0) + \underline{w}_0)^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(x-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx & (\text{i}^{***}) \\
&+ \int_{\underline{w}_0}^{\overline{w}_0} x^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(x-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx & (\text{ii}^{***}) \\
&+ \int_{\overline{w}_0}^{+\infty} (\gamma \cdot (x - \overline{w}_0) + \overline{w}_0)^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(x-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx & (\text{iii}^{***}) \\
&+ \int_0^{\underline{w}_0} (\alpha \cdot (x - \underline{w}_0) + \underline{w}_0)^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((-x)-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx & (\text{iv}^{***}) \\
&+ \int_{\frac{\overline{w}_0 - \underline{w}_0}{\alpha} + \underline{w}_0}^{\underline{w}_0} (\alpha \cdot (x - \underline{w}_0) + \underline{w}_0)^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((-x)-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx & (\text{v}^{***}) \\
&+ \int_{\frac{\overline{w}_0 - \underline{w}_0}{\alpha} + \underline{w}_0}^{+\infty} (\alpha \cdot (x - \underline{w}_0) + \underline{w}_0)^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((-x)-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx & (\text{vi}^{***}) \\
&+ \int_0^{\frac{\underline{w}_0}{\alpha} - \underline{w}_0} (\alpha \cdot ((-x) - \underline{w}_0) + \underline{w}_0)^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{((-x)-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx & (\text{vii}^{***}) \\
&+ \int_{\frac{\underline{w}_0}{\alpha} - \underline{w}_0}^{\underline{w}_0} (\alpha \cdot ((-x) - \underline{w}_0) + \underline{w}_0)^\eta \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 w^2)}} e^{-\frac{(x-\beta w)^2}{2 \cdot (\alpha_0 + \alpha_1 w^2)}} dx & (\text{viii}^{***})
\end{aligned}$$

- (a) Der einfachste Vergleich ist nun zwischen (ii) und (ii<sup>\*\*\*</sup>) möglich, da diese Integrale übereinstimmen und somit auf die Erwartungswertbildung den gleichen Effekt haben.
- (b) Betrachtet man die Summen [(v) + (vi)] und [(v<sup>\*\*\*</sup>) + (vi<sup>\*\*\*</sup>)], so hat man es in beiden Fällen mit einem Integral von  $\underline{w}_0$  bis  $+\infty$  zu tun. Für  $x > \underline{w}_0$  gilt entsprechend der Glättungs-idee  $(\alpha(x - \underline{w}_0) + \underline{w}_0) < x$  und somit  $[(v^{***}) + (vi^{***})] < [(v) + (vi)]$  mit Monotonie von „hoch“  $\eta$ .
- Im Sinne der Verschiebung der Dichte, bedeutet dies eine Verschiebung nach links in Richtung  $\underline{w}_0$ , da die Dichtefunktion  $f(x)$  an der Stelle  $x$ , eben nicht mehr mit  $x$ , sondern mit einem kleineren Wert multipliziert wird (hervorgerufen durch die Glättung). Der Grund für dieses Phänomen liegt darin, dass durch die Glättungsfunktion die Werte kleiner als  $(-\underline{w}_0)$  unwahrscheinlicher werden und durch die Betragsfunktion dann die Werte größer  $\underline{w}_0$ . Die Ausdrücke (iv), (v) und (vi) spiegeln ja die Wahrscheinlichkeitsmasse wider, die zunächst auf  $\mathbb{R}_-$  entfallen war.
- (c) Die Integrale (iii) und (iii<sup>\*\*\*</sup>) haben den gleichen Integrationsbereich und da für  $x > \overline{w}_0$  wiederum gilt:  $(\gamma(x - \overline{w}_0) + \overline{w}_0) < x$ , folgt mit Monotonie von „hoch“  $\eta$  und den gleichen

Überlegungen wie in (b) auch (iii<sup>\*\*\*</sup>) < (iii), sowie die Interpretation in Hinblick auf die Verschiebung der Dichte.

(d) Die Integrale (iv) und (iv<sup>\*\*\*</sup>) haben den gleichen Integrationsbereich und da für  $x \in (0, \underline{w}_0)$  entsprechend der Glättungsidee  $(\alpha(x - \underline{w}_0) + \underline{w}_0) > x$  gilt, folgt (iv<sup>\*\*\*</sup>) > (iv) mit Monotonie von „hoch“  $\eta$ . Quasi spiegelverkehrt zu den Überlegungen von (b),(c) wird die Dichte  $f(x)$  hier nicht mit kleineren, sondern mit größeren Werten als  $x$  multipliziert - d.h. es findet eine Verschiebung der Dichte nach rechts in Richtung  $\underline{w}_0$  statt.

(e) Auch (i) und (i<sup>\*\*\*</sup>) führen mit  $(\alpha(x - \underline{w}_0) + \underline{w}_0) > x$  für  $x \in (0, \underline{w}_0)$  und den gleichen Überlegungen wie in (d) zu (i<sup>\*\*\*</sup>) > (i).

(f) Die Integrale (vii<sup>\*\*\*</sup>) und (iix<sup>\*\*\*</sup>) haben zwar keine Entsprechung im ungeglätteten Fall, entspringen jedoch (i<sup>\*\*\*</sup>) und (iv<sup>\*\*\*</sup>). Der Integrationsbereich ist mit  $(0, \underline{w}_0/\alpha - \underline{w}_0)$  abhängig von der Wahl von  $\alpha$  und kann also beliebig groß werden.

Die Interpretation dieses Integrationsbereichs ist die, dass damit der Bereich abgedeckt ist über dem die Dichte im ungeglätteten Fall betrachtet würde. Durch das Glätten wird die Dichte dann aber eben nicht auf diesen Bereich entfallen, sondern auf den Bildbereich von  $(\alpha \cdot ((-x) - \underline{w}_0) + \underline{w}_0)^\eta$ , der gegeben ist durch  $(0, \alpha \cdot (-\underline{w}_0) + \underline{w}_0) \subset (0, \underline{w}_0)$  und somit genau die Lücke schließt, die in (i<sup>\*\*\*</sup>) bzw. (iv<sup>\*\*\*</sup>) beim Bildbereich des Vorfaktors besteht. Da für  $x \in (0, \underline{w}_0/\alpha - \underline{w}_0)$ , der Vorfaktor auf  $(0, \alpha \cdot (-\underline{w}_0) + \underline{w}_0)$  abbildet und somit auf diesen Bereich zusätzlich Wahrscheinlichkeitsmasse entfällt, hat das auf die Erwartungswertbildung den Effekt, dass eine Verschiebung nach unten in Richtung Nullpunkt geschieht.

Zusammengenommen liefern (a), (b) und (c) bzw. (ii<sup>\*\*\*</sup>), (v<sup>\*\*\*</sup>), (vi<sup>\*\*\*</sup>) und (iii<sup>\*\*\*</sup>) ein Bild über den Bereich  $(\underline{w}_0, +\infty)$ . Die Dichte in diesem Bereich wird jeweils mit einem kleineren Vorfaktor multipliziert, was ja wie oben gesehen als Verschiebung in Richtung Kompaktum oder in Richtung  $\underline{w}_0$  interpretiert werden kann. In Hinblick auf den Erwartungswert ist die Sache so zu interpretieren, dass besonders große Werte damit unwahrscheinlicher werden und der Erwartungswert damit in Richtung Kompaktum rückt!

Des Weiteren liefern (d) und (e) bzw. (i<sup>\*\*\*</sup>) und (iv<sup>\*\*\*</sup>), dass die Dichte die vorher auf  $(0, \underline{w}_0)$  entfallen war, durch das Glätten auf  $(\alpha \cdot (-\underline{w}_0) + \underline{w}_0, \underline{w}_0)$  entfällt oder anders ausgedrückt in Richtung  $\underline{w}_0$  verschoben wird. In Hinblick auf den Erwartungswert, bedeutet dies natürlich erst einmal eine Steigerung, die aber durch (f) zumindest teilweise relativiert wird.

Mit (f) sieht man, dass zusätzliche Wahrscheinlichkeitsmasse auf den Bereich  $(0, \alpha \cdot (-\underline{w}_0) + \underline{w}_0)$  entfällt und den Erwartungswert damit sogar noch in Richtung 0 zieht oder kleiner werden lässt.

Einzig und allein die Tatsache, dass mit (d) und (e) eine Verschiebung der Wahrscheinlichkeitsmasse nach rechts oder nach oben in Richtung Kompaktum geschieht, behindert einen darin die gewünschten Schlüsse zu ziehen, da diese Verschiebung ja insbesondere auch von der Wahl von  $\alpha$  abhängt. Wir nehmen an dieser Stelle einfach den schlimmsten Fall an und wollen auf dieser Grundlage schlussfolgern. Mal angenommen die gesamte Wahrscheinlichkeitsmasse, die im ungeglätteten Fall auf  $(0, \underline{w}_0)$  entfallen war, würde im geglätteten Fall komplett auf  $\underline{w}_0$  liegen, hätte das eine Verschiebung des Erwartungswerts nach oben zur Folge und zwar im

schlimmsten Fall um eine Konstante vom Betrag  $\underline{w}_0$ . Da die Verschiebung von der Wahl von  $\alpha$  abhängt und in Richtung  $\underline{w}_0$  geht, sei die Konstante bezeichnet mit  $\underline{w}_0^\alpha$ .

Die Aussage für den ungeglätteten Prozess lautete ja, dass für  $w \in A^c$ ;  $A = (-M, M)$  sich ergibt

$$1 + (6) \leq (1 - \delta)h(w).$$

Um die Aussage auf (7) zu übertragen, lassen sich die bisherigen Ergebnisse wie folgt umsetzen:

$$1 + (7) \leq 1 + (6) + \underline{w}_0^\alpha \leq (1 - \delta)h(w) + \underline{w}_0^\alpha = \left[ (1 - \delta) + \frac{\underline{w}_0^\alpha}{1 + |w|^\eta} \right] \cdot (1 + |w|^\eta)$$

mit  $h(w) = 1 + |w|^\eta$ . Wählt man nun  $M$  hinreichend groß (siehe Definition von  $A^c$ ), damit es  $\frac{\underline{w}_0^\alpha}{1 + |M|^\eta} < \delta$  erfüllt, so folgt für alle  $w \in A^c$

$$1 + (7) \leq \left[ (1 - \delta) + \frac{\underline{w}_0^\alpha}{1 + |w|^\eta} \right] \cdot (1 + |w|^\eta) \leq \left[ (1 - \delta) + \frac{\underline{w}_0^\alpha}{1 + |M|^\eta} \right] \cdot (1 + |w|^\eta) \leq (1 - \delta^*)h(w)$$

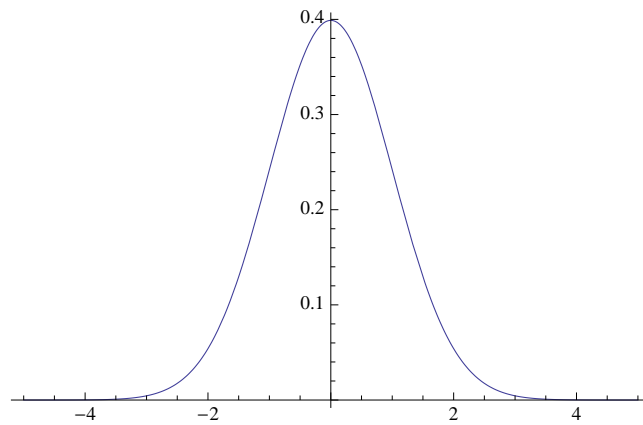
bei der Festlegung  $\delta^* = (\delta - \frac{\underline{w}_0^\alpha}{1 + |M|^\eta})$ .

Dieses Ergebnis vervollständigt die Argumentation im Beweis zu Satz 2.7 und bedeutet, zumindest Eingeschränkt auf das AR(1)-ARCH(1)-Modell, dass sich die Driftbedingung vom ungeglätteten zum geglätteten Prozess übertragen lässt.

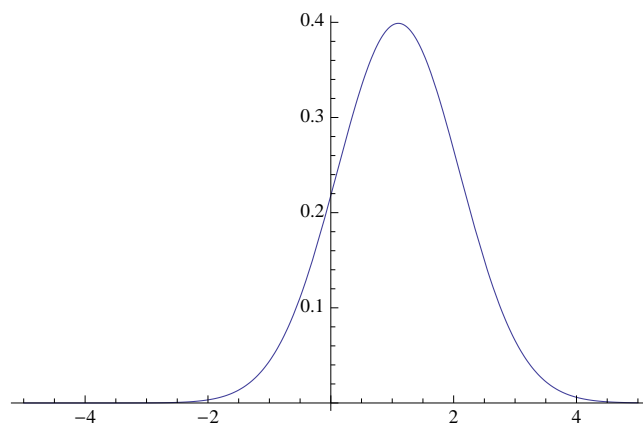
### 4.3 Zusammenhang von $\beta$ und $\alpha_1$

In Abschnitt 2 und 3 war stets die Beziehung zwischen dem ungeglätteten und geglätteten Fall ein zentraler Ausgangspunkt für die Betrachtung. Im Detail standen zum Beispiel die zulässigen Parameterwerte für die jeweils stationären Lösungen zur Diskussion. Da die tabellierten Werten für  $\alpha_1$  und  $\beta$  in BuK [19] wie gesehen nur durch ein Iterationsverfahren erzeugt werden (vgl. Abschnitt 2), können Zweifel aufkommen an ihrer Richtigkeit. Indem man den Ausdruck  $\log |\beta + \sqrt{\alpha_1}\varepsilon|$  genauer betrachtet, kann man jedoch eine Plausibilitätsüberlegung und Veranschaulichung anstellen, um so die Entstehung der Werte nachzuvollziehen. Von  $\varepsilon$  wurde ja gefordert, dass es symmetrisch um den Nullpunkt ist und eine positive Dichte über  $\mathbb{R}$  hat. Im vorliegenden Beispiel wird die Standard-Normalverteilung angenommen und diese besitzt nicht nur die geforderten Eigenschaften, sondern konzentriert darüber hinaus auch einen Großteil der Masse um den Nullpunkt. Dieser Umstand hat zur Folge, dass bei einem  $\beta$ , dessen Betrag nur geringfügig  $> 1$  ist, einfach wegen der Symmetrie und mit hoher Wahrscheinlichkeit, der Gesamtausdruck  $|\beta + \sqrt{\alpha_1}\varepsilon|$  Werte  $< 1$  hervorbringt.

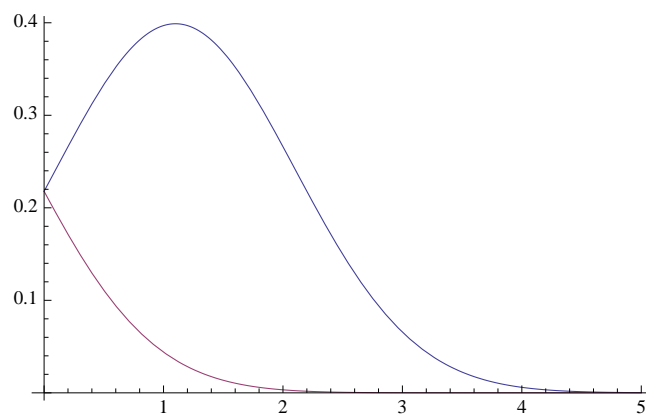
Ganz anschaulich kann man sich dies anhand der für die Normalverteilung typischen Glockenkurve verdeutlichen. Die Zufallsgröße  $\varepsilon$  folgt der Standardnormalverteilung und hat somit eine Dichtefunktion von der Gestalt:



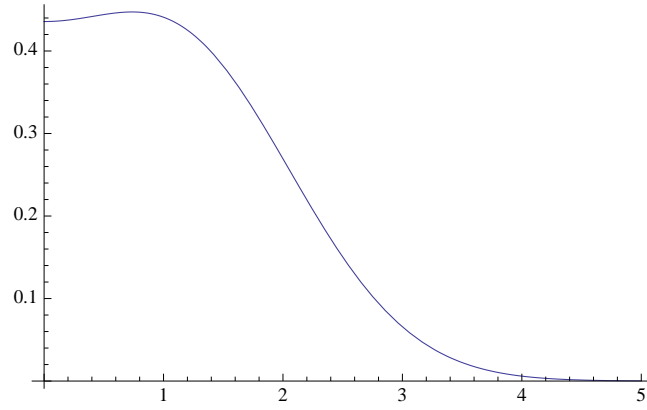
Eine Addition von  $\beta$  hat eine Verschiebung zur Folge und eine hier nicht abgebildete Multiplikation mit  $\sqrt{\alpha_1}$ , würde eine Streckung oder Stauchung bedeuten.



Die Betragsstriche haben den Effekt, dass die Wahrscheinlichkeitsmasse, die zuvor den negativen Werten zugeordnet wurde, auf die positiven Werte aufgeschlagen wird (dies entspricht einer Spiegelung an der y-Achse).



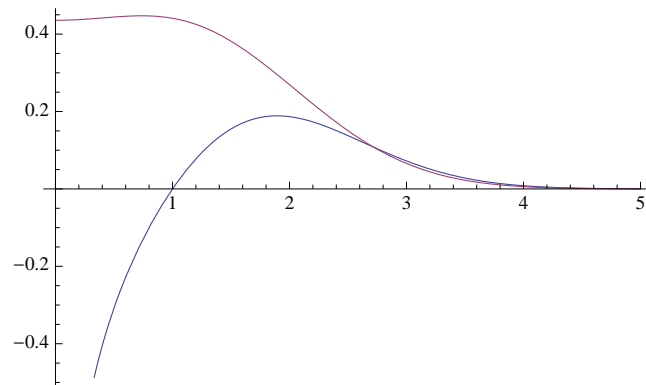
Insgesamt ergibt sich also die folgende Dichte nach Anwendung der Betragsfunktion:



Die obigen Abbildungen behandeln den Fall  $\beta = 1.1$  und  $\alpha_1 = 1$ , der ja entsprechend der Tabellen auch  $h'_{\beta, \alpha_1}(0) = \mathbb{E}(\log |\beta + \sqrt{\alpha_1} \varepsilon|) < 0$  erfüllt.

Wenn man sich die letzte Abbildung ansieht, so erkennt man unschwer, dass auf den Wertebereich  $0 < x < 1$  ein nicht unwesentlicher Teil der Wahrscheinlichkeitsmasse entfällt. Bei der Erwartungswertbildung fallen diese Werte nun besonders schwer ins Gewicht und mit dem Logarithmus auch mit deutlich negativem Vorzeichen. Diese Überlegung spiegeln die Ergebnisse der Tabellen wider: Wegen der Symmetrie ergibt sich bei einem Wert von  $|\beta|$ , der zumindest nicht wesentlich über 1 liegt, ganz automatisch der Umstand, dass ein Großteil der Wahrscheinlichkeitsmasse sich auf die Werte  $0 < x < 1$  verteilt. Bei Werten deutlich über 1, verschiebt sich auch der Großteil der Wahrscheinlichkeitsmasse und der Logarithmus kommt nicht mehr so stark zum tragen. Eine konsequente Fortführung dieses Gedankenganges erklärt auch die Werte für  $\alpha_1$  und den Zusammenhang zu  $\beta$ . Bei größer werdendem  $|\beta|$ , verkleinert sich der zulässige Bereich für  $\alpha_1$ , was nur logisch erscheint, da durch eine Verschiebung immer mehr Wahrscheinlichkeitsmasse auf Werte  $> 1$  entfällt. Durch ein kleiner gewähltes  $\alpha_1$ , konzentriert sich die Masse mehr um den Erwartungswert und somit auch mehr auf den für unsere Überlegung interessanten Bereich  $0 < x < 1$ . Man kann also im gewissen Maße die Verschiebung durch ein kleiner werdendes  $\alpha_1$  kompensieren. Ist der Wert  $|\beta| = 1$  überschritten, so kommt noch hinzu, dass ein gewisses Maß an Schwankung nötig ist, um den interessanten Bereich überhaupt noch zu erreichen und somit schrumpft der zulässige Bereich für  $\alpha_1$ , ab diesem Punkt auch von links (von Null) kommend.

All diese Überlegungen und Abbildungen können den Zusammenhang und zulässigen Wertebereich von  $\beta$  und  $\alpha_1$  ganz hervorragend beschreiben und plausibel machen, aber die eigentlichen Zahlenwerte sind dadurch noch nicht gewonnen.

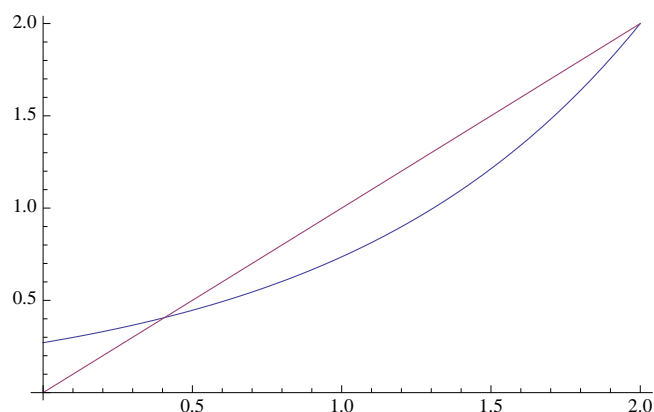


Als Abschluss der obigen Überlegungen, zeigt die letzte Graphik zum einen nochmals die Verteilungsdichte (ergibt beim Integrieren den Wert 1) und zum anderen den Graphen des Ausdrucks  $\mathbb{E}[\log |\beta + \sqrt{\alpha_1} \varepsilon_t|]$ . Auch wenn die konkreten Werte dadurch nicht berechnet wurden, kann man doch zweifelsfrei die Gesetzmässigkeit erkennen, die ihrer Entstehung zugrunde liegt.

#### 4.4 Glättungsfunktion am Rand

Im Verlauf der Arbeit wurden mehrere Glättungsfunktionen behandelt und ihre Eigenschaften in Hinblick auf die Erfüllung einer Kontraktionsbedingung variiert bzw. angepasst - siehe z.B. Abschnitt 2.1. Ein besseres Verständnis der gemachten Forderungen und ihrer Konsequenzen, soll im Folgenden durch eine Veranschaulichung und die Verwendung eines konkreten Zahlenbeispiels unterstützt werden. Die Geradenglättung als unmittelbare Konsequenz aus diesen Überlegungen fand sich ja in mehreren Beweisen und grundsätzlichen Überlegungen wieder.

Um sich in die Problematik des Glättens am Rand besser hineinendenken zu können, ist eine konkrete Wahl des Kompaktums und eine graphische Veranschaulichung sicherlich förderlich. Man wähle daher  $\underline{w}_0 = 2$  und  $\overline{w}_0 = 5$ .



Für diese Konstellation bedeutet Glätten unterhalb von  $\underline{w}_0 = 2$  dann ganz anschaulich folgendes: Alle Zahlen, die kleiner sind als 2, werden auf den Bereich zwischen 0 und 2 abgebildet, jedoch bildet beispielsweise  $\underline{w}_0 e^{-(\underline{w}_0 - w')}$  (blau) als Kandidat für eine Glättungsfunktion erst eine Kontraktion, wenn die Ableitung kleiner 1 ist. Eine Abbildung  $f$  ist nur dann eine Kontraktion, wenn insbesondere  $|f(x) - f(y)| < |x - y|$  gilt (s.o.). Die rote Gerade bildet die Identität ab

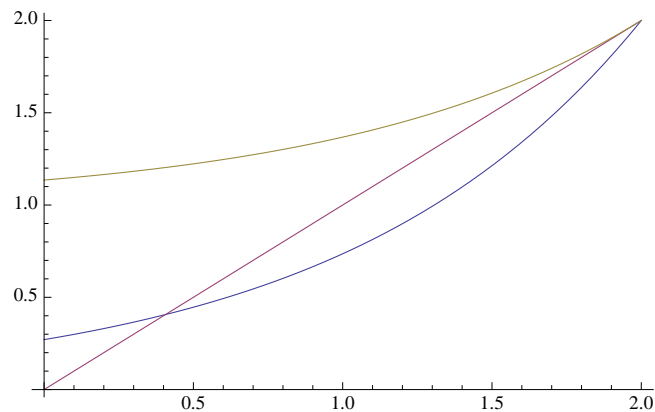
und hilft das Verhalten der Glättungsfunktion in blau zu verstehen.

Aus der Ableitung von  $g(w') = \underline{w}_0 e^{-(\underline{w}_0 - w')}$  erhält man den Bereich

$$w' < \ln\left(\frac{1}{\underline{w}_0}\right) + \underline{w}_0 \quad (\approx 1.3069 \text{ für } \underline{w}_0 = 2),$$

als den kontraktiven. Für den Gesamtprozess bedeutet dies bis hierher, dass das Verhalten im Bereich der Nahtstelle  $(1.3069, 2)$  eben nicht der gewünschten Kontraktionsbedingung genügt und somit für das Verhalten des Prozesses gesondert untersucht werden müsste.

Um diesem Problem aus dem Weg zu gehen, verstärken wir die Eigenschaft der Glättungsfunktion und stauchen den Prozess noch etwas stärker.

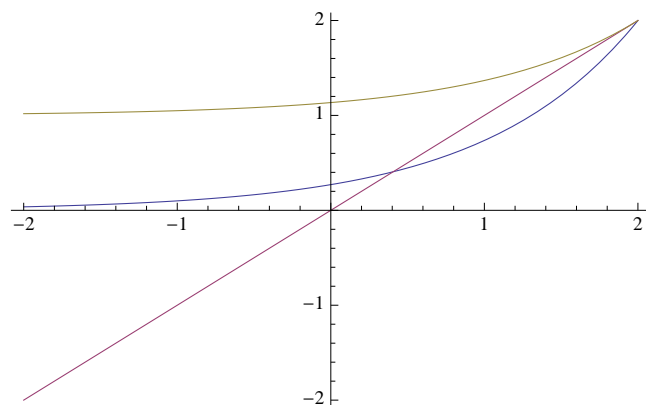


Die Glättungsfunktion wird dahingehend verändert, dass sie zwar jede Zahl  $< \underline{w}_0$  auf den Bereich  $(0, \underline{w}_0)$  abbildet, aber immer eine Steigung  $< 1$  besitzt. Dies geschieht durch die Festlegung auf:

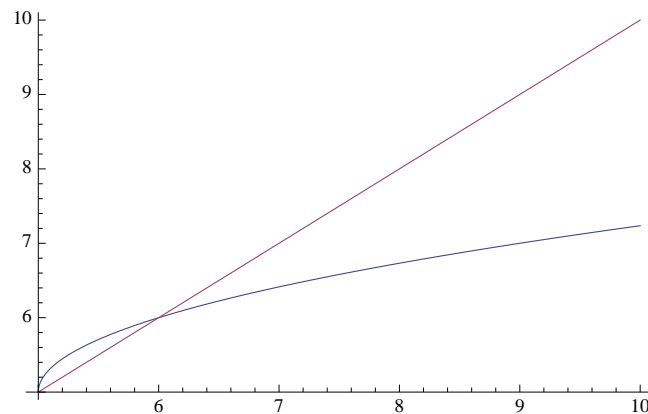
$$g^*(w') = e^{-(\underline{w}_0 - w')} + (\underline{w}_0 - 1),$$

für  $w' < \underline{w}_0$  und  $\underline{w}_0 \geq 1$ . Die Einschränkung auf  $\underline{w}_0 \geq 1$  erscheint dabei verschmerzbar zu sein und bietet, wenn man sich an das Aktienbeispiel erinnert, wenig Plausibilitätsprobleme. Die gelb eingezeichnete neue Glättungsfunktion ist also eine deutlich stärkere Forderung und besonders der Umstand, dass sie über der Identitätsgeraden liegt, wird sich als nützlich erweisen und soll deshalb hier bemerkt werden.

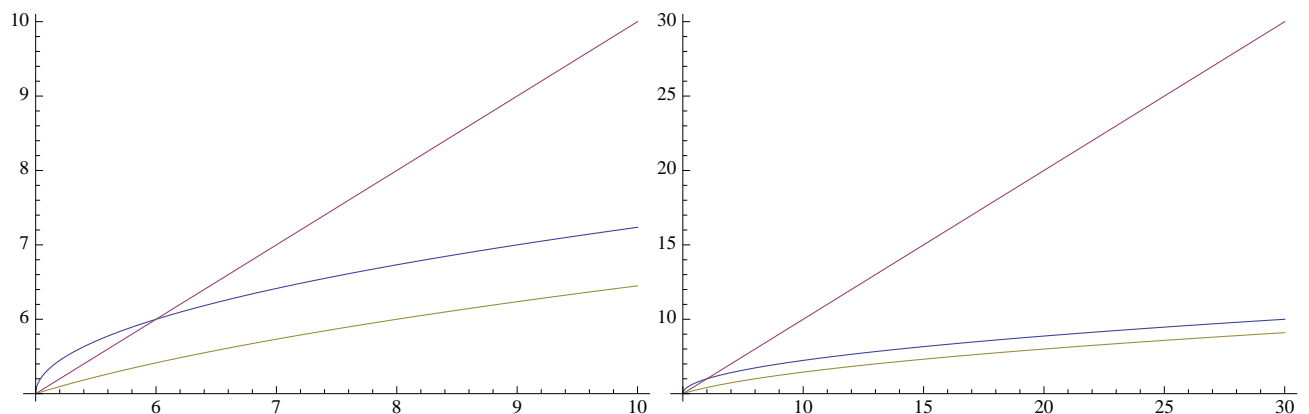
Natürlich ist das Verhalten nicht nur auf  $\mathbb{R}^+$  eingeschränkt, sondern zieht sich ins Negative fort.



Am anderen Ende des Kompaktums begegnet einem ein ganz analoges Problem. Wiederum zur besseren Darstellung sei nun  $\overline{w}_0 = 5$  und  $\alpha = \frac{1}{2}$  gewählt. Für Werte größer 5 bedeutet glätten demnach: Ein Ausbrechen des Prozesses wird verhindert durch eine Stauchung von oben, jedoch geschieht dies nicht so drastisch, wie am unteren Ende des Kompaktums. Die Zahl 0 stellte für die Abweichung nach unten eine Schranke dar, deren Bedeutung ja bereits in der Modellkonstruktion begründet war. Eine derartige Schranke auch nach oben hin anzugeben erscheint da weniger sinnvoll, insbesondere vor dem Hintergrund eines Aktienbeispiels, dem man damit eine Wachstumsgrenze auferlegen würde. Nichts desto trotz war ja aus den einführenden Überlegungen heraus ein  $\overline{w}_0$  als gute Schranke nach oben gewählt worden und eine Überschreitung  $w' > \overline{w}_0$  soll durch den Glättungskandidaten  $g(w') = \overline{w}_0 + (w' - \overline{w}_0)^\alpha$  (blau) zumindest teilweise abgefangen werden.



Ganz analog zu der Situation am unteren Ende des Kompaktums, ergibt sich hier im Bereich von  $(\overline{w}_0, \alpha^{\frac{1}{1-\alpha}} + \overline{w}_0)$  ein Fenster, in dem die Glättungsfunktion eben keine Kontraktion darstellt. Für das konkrete Beispiel bedeutet dies, dass  $(5, 5\frac{1}{4})$  einer gesonderten Betrachtung bedürfe. Man umgeht dies Problem wiederum durch die Festlegung auf eine stärkere Glättungsfunktion



$g^*(w') = (\overline{w}_0 - 1) + (w' - \overline{w}_0 + 1)^\alpha$  (gelb) für  $w' > \overline{w}_0$ . Die Verschiebung um den Wert 1 rührt von daher, dass  $0 < \alpha < 1$  auch gleich  $0 < \alpha^{\frac{1}{1-\alpha}} < 1$  nach sich zieht und der nicht-kontraktive Bereich somit nicht weiter als zu  $\overline{w}_0 + 1$  reicht. Insgesamt bietet die neue stärkere Glättungsfunktion



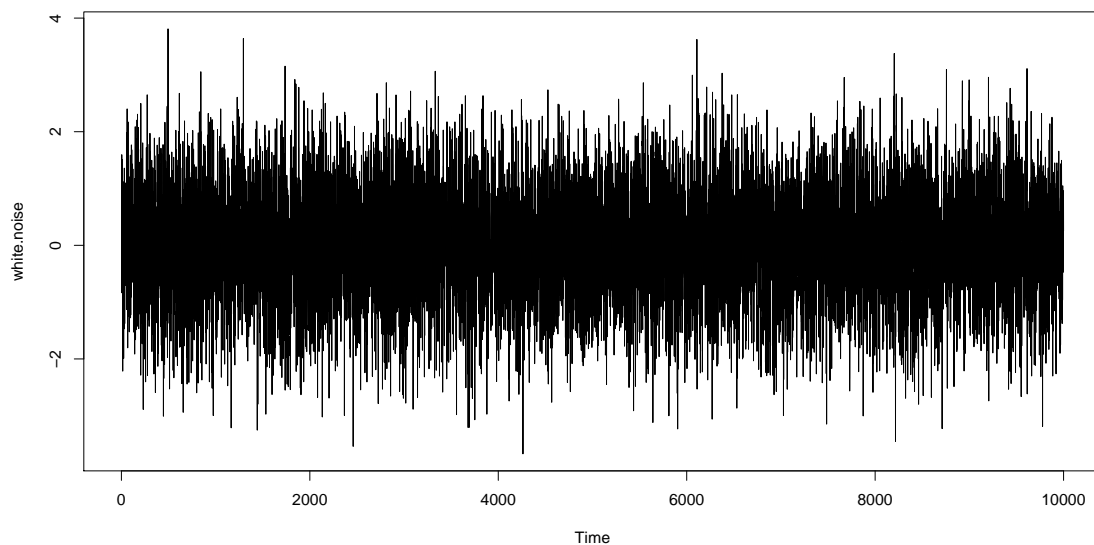
$$g^*(w') = \begin{cases} e^{-(\underline{w}_0 - w')} + (\underline{w}_0 - 1) & , w' < \underline{w}_0, \underline{w}_0 \geq 1 \\ w' & , \underline{w}_0 \leq w' \leq \overline{w}_0 \\ (\overline{w}_0 - 1) + (w' - \overline{w}_0 + 1)^\alpha & , \overline{w}_0 < w' \quad 0 < \alpha < 1 \end{cases}$$

also die gewünschten Eigenschaften. Ganz anschaulich bedeutet dies für den dadurch entstehenden Gesamtprozess folgendes: Wenn der Ausgangsprozess sich in dem durch die Werte  $\underline{w}_0$  und  $\overline{w}_0$  festgelegten Bereich bewegt, so hat „smoothing / Glätten“ natürlich keinen Effekt und der Gesamtprozess unterscheidet sich nicht vom Ausgangsprozess. Wird dieses Kompaktum  $[\underline{w}_0, \overline{w}_0]$  jedoch vom Ausgangsprozess verlassen, so wird jeder Wert  $w'$  verändert zu einem Wert  $\tilde{w}'$ , der zwar weiterhin außerhalb des Kompaktums liegt, aber deutlich näher an das Kompaktum rückt (der veränderte Wert geht in die nächste Iteration ein und beeinflusst die weiteren Werte, aber die Berechnungsvorschrift bleibt erhalten). Unterhalb von  $\underline{w}_0$  findet eine Verschiebung nach oben und oberhalb von  $\overline{w}_0$  eine Verschiebung nach unten statt. Diese Eigenschaft erhält man erst durch die neue Glättungsfunktion  $g^*$  und somit eine Stauchung in Richtung Kompaktum.

## 4.5 Prozess-Bausteine

Einige sehr weit verbreitete stochastische Prozesse sind dadurch entstanden, dass bereits bestehende Prozesse aufgegriffen und in Hinblick auf fehlende Eigenschaften erweitert oder abgeändert wurden. Die Entstehungsgeschichte eines stochastischen Prozesses ist in der Anwendung natürlich nicht von Belang, jedoch für ein tieferes Verständnis kann es von immensem Vorteil sein nicht nur das Gesamtobjekt, sondern auch seine Zusammensetzung, zu kennen. Insbesondere vor dem Hintergrund der computerunterstützten Umsetzung, ist es immer wieder schon auf Grund der übersichtlichen Darstellung unausweichlich, dem Gesamtprogramm eine sinnvolle Struktur zu geben. Eben der Prozessentstehungsgeschichte folgend, ergeben sich mehr oder weniger natürlich einzelne Teilprogramme.

Der White-Noise-Prozess  $(\varepsilon_t, t \in \mathbb{N})$  (z.B. [37] S.10) entsteht aus einer i.i.d.-Folge von Zufallsvariablen, die zumeist mit Erwartungswert 0 und Varianz 1 ausgestattet sind. Für sich genommen bietet der Prozess wegen mangelnder Flexibilität und Entsprechung in der Realität wenig Potenzial. Als Störterm oder unbeobachtbarer Zufallseinfluss, spielt er in der Konstruktion weit verbreiteter stochastischer Prozesse jedoch eine entscheidende Rolle. Der klassische Fall wird über die  $\mathcal{N}(0,1)$ -Verteilung erzeugt. Die Namensgebung White-Noise oder Weisses Rauschen erklärt sich am besten anhand eines typischen Schaubildes.



Eine White-Noise Simulation in R, kann wie folgt aussehen:

```
# Startwert für den Zufallsgenerator festlegen
> set.seed(1)
# 10.000 Zufallszahlen aus der Standardnormalverteilung erzeugen
> eps<-rnorm(10000)
# Erzeugte Werte in time-series-Element speichern
> eps<-ts(eps)
```

Zum Verständnis des Programmcodes gibt es dabei zu sagen, dass die mit # eingeleiteten Zeilen als Kommentar unberücksichtigt bleiben, das Symbol > jeweils eine neue Befehlszeile einleitet und die obige Grafik sich mit dem Befehl

```
> plot(eps)
```

erzeugen lässt.

Diese recht übersichtliche Darstellung und die freie Zugänglichkeit des Softwarepaketes R<sup>19</sup>, machen es zur ersten Wahl für die angestrebten Implementierungen, da alle Ergebnisse somit für Jedermann mit wenig Aufwand und vor allem kostenlos reproduzierbar und überprüfbar sind. Solch eine Transparenz kann mit Matlab, Maple oder anderer Software nicht erreicht werden. Einen guten Einstieg in den Umgang mit R, liefert Jürgen Groß [10] und bietet auch die Grundlage für den Großteil der hier eingebundenen Grafiken und Ergebnisse. Eine White-Noise-Simulation findet man beispielsweise unter dem Begriff „Weisses Rauschen“ in Kapitel 23.

Ein autoregressiver Prozess trägt seinen Namen bekanntlich deshalb, weil sein Wert zum Zeitpunkt  $t$  unmittelbar von den vorangegangenen Werten  $t - 1, \dots$  abhängt. Der funktionale

---

<sup>19</sup><http://www.r-project.org/>

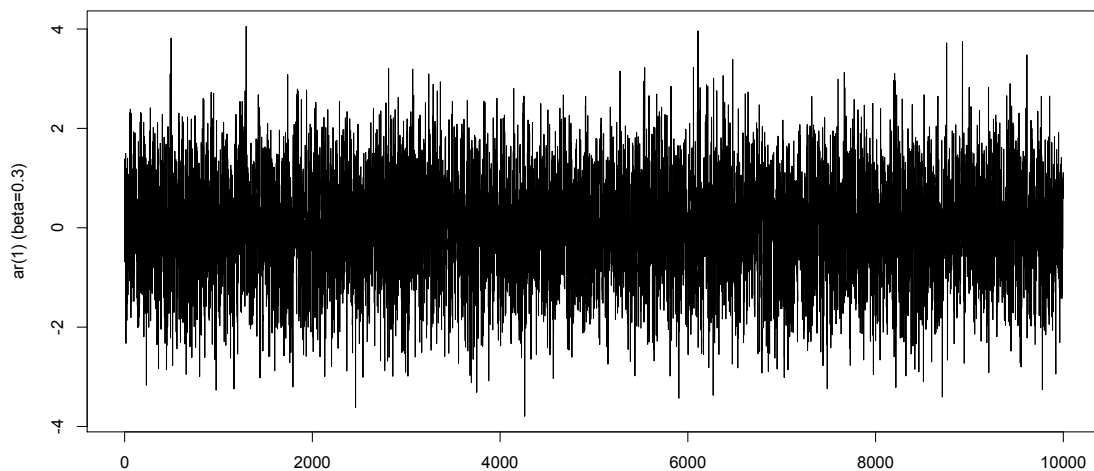
Zusammenhang kann je nach Definition variieren und der Vergangenheitsbezug ebenso, aber naiv gesprochen findet eine Regression über die vergangenen Werte statt. Der oben simulierte White-Noise-Prozesses fließt, wie bereits erwähnt, als Störterm in die Konstruktion des AR(1)-Prozesses ein und wird an den autoregressiven Part angehängt/addiert (vgl. [10],[37],[39]). Da ein AR-Prozess eben keine Folge von i.i.d. Zufallsvariablen ist, sondern einen Vergangenheitsbezug bzw. eine Rekursion aufweist, bedarf es zu seiner Simulation einer Schleife - (FOR-Schleife).

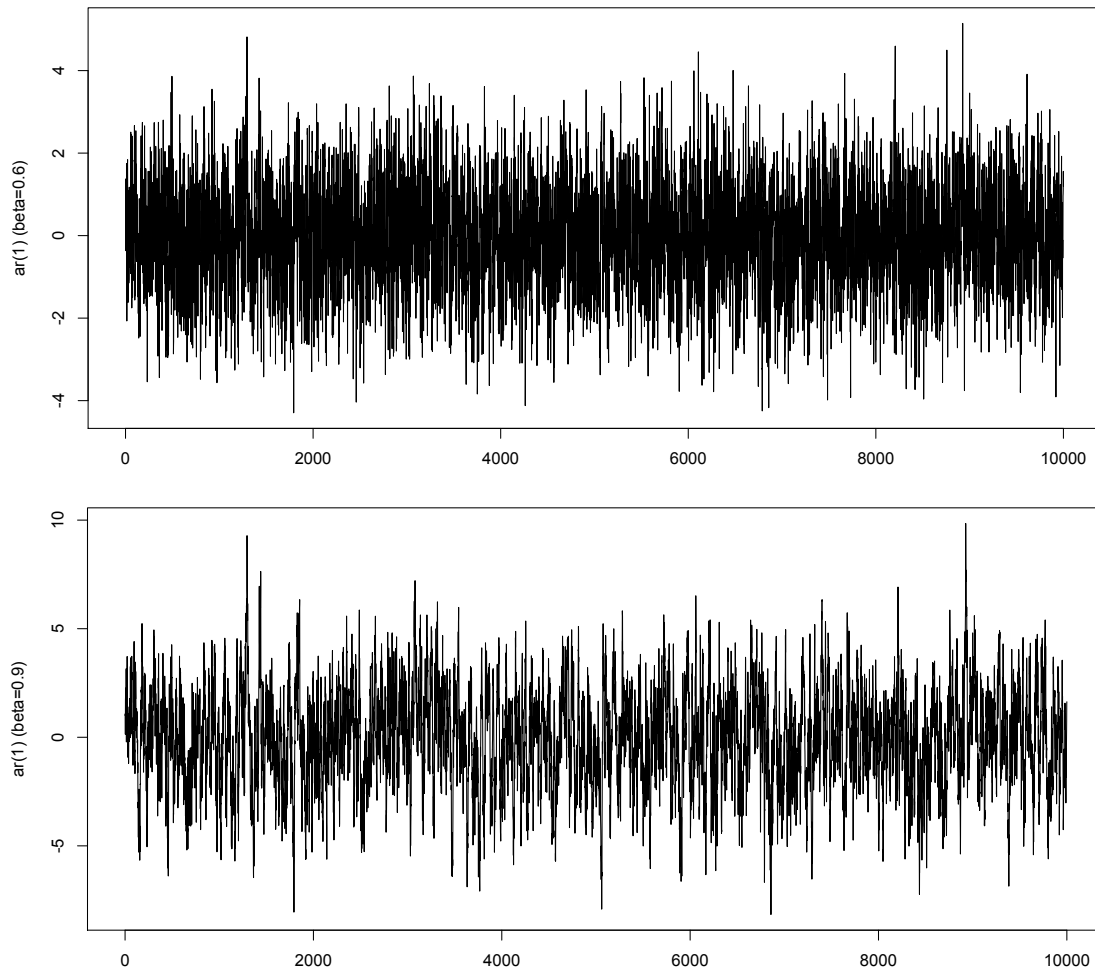
#### Simulation eines AR(1)-Prozesses

$$W_{t+1} = \beta W_t + \varepsilon_{t+1}$$

```
# Erzeuge einen Vektor mit den Zahlen 1 bis 10000 als Platzhalter bzw. Startwert
> w<-c(1:10000)
# Lege den Parameter  $\beta$  fest
> beta<-0.9
# Erzeuge aus dem Vektor ein time-series-Element
> w<-ts(w)
# Rekursiv wird der AR(1)-Prozess erzeugt
> for(i in 1:9999){
+ w[i+1]<-beta*w[i] + eps[i+1]}
> ar1<-w
```

Das einleitende + in der vorletzten Zeile deutet hierbei an, dass die vorangegangene Zeile hier weiter geht und hat sonst keine weitere Bedeutung. Durch den Parameter  $\beta$  besitzt dieser Prozesstyp, trotz des in unserem Fall eingeschränkten Vergangenheitsbezugs, eine starke Variabilität.





Obige Schaubilder zeigen die Ergebnisse für den Parameter  $\beta = 0.3, 0.6$ , sowie  $0.9$  und die damit einhergehende Veränderung. Der Störterm ist in den drei Schaubildern identisch und alle Unterschiede damit zurückführbar auf den Parameter  $\beta$ . Andererseits sind einige besonders hohe Ausschläge stets an gleicher Stelle erkennbar und verdeutlichen wie ein solches Schaubild im Zusammenspiel von AR und Störterm entsteht.

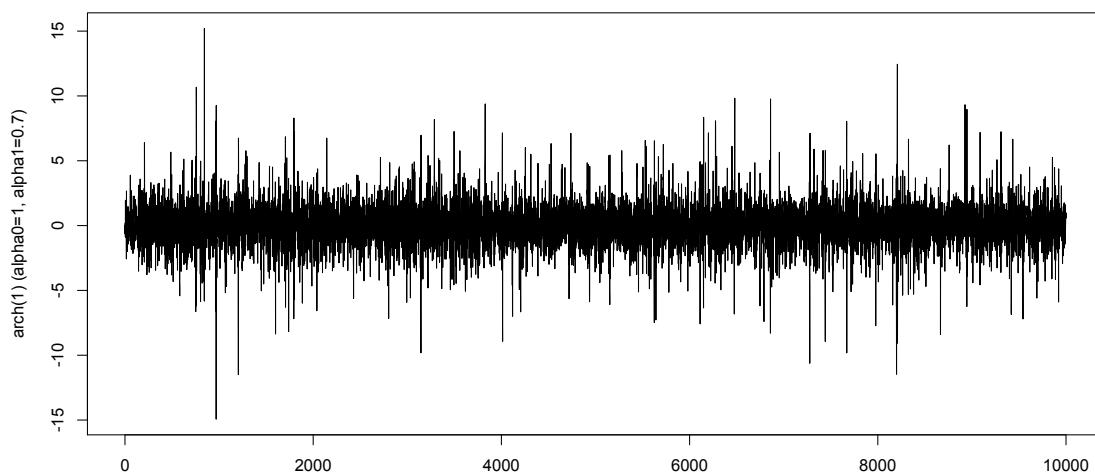
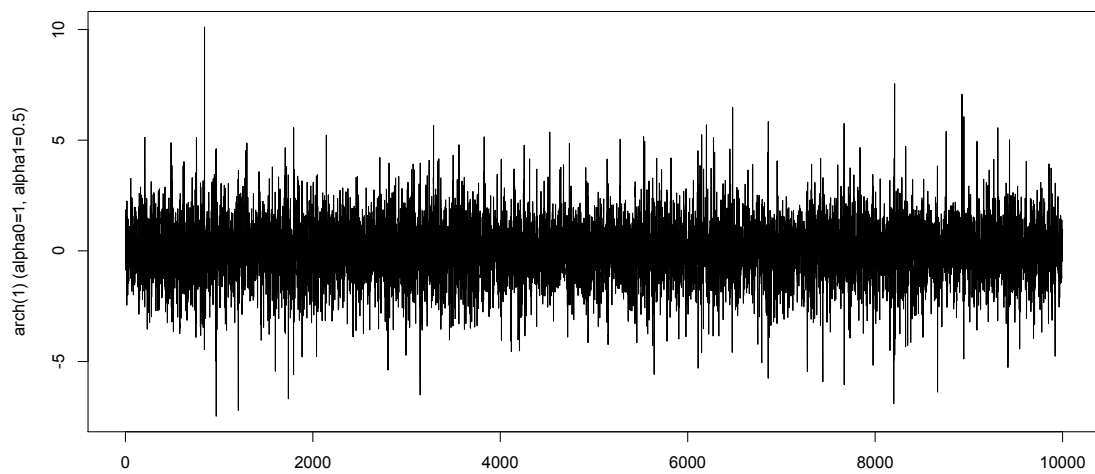
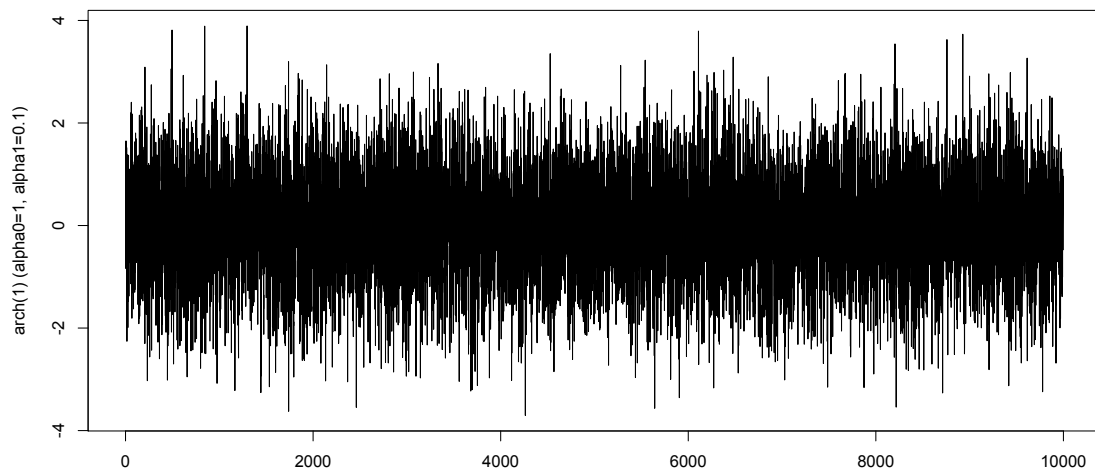
Ein ARCH(1)-Prozess (vgl. [37] S.17) ist ein weiterer wichtiger Baustein für die in dieser Arbeit auftauchenden Prozesse. Gemäß der oben verwendeten Notation für AR(1)-ARCH(1)

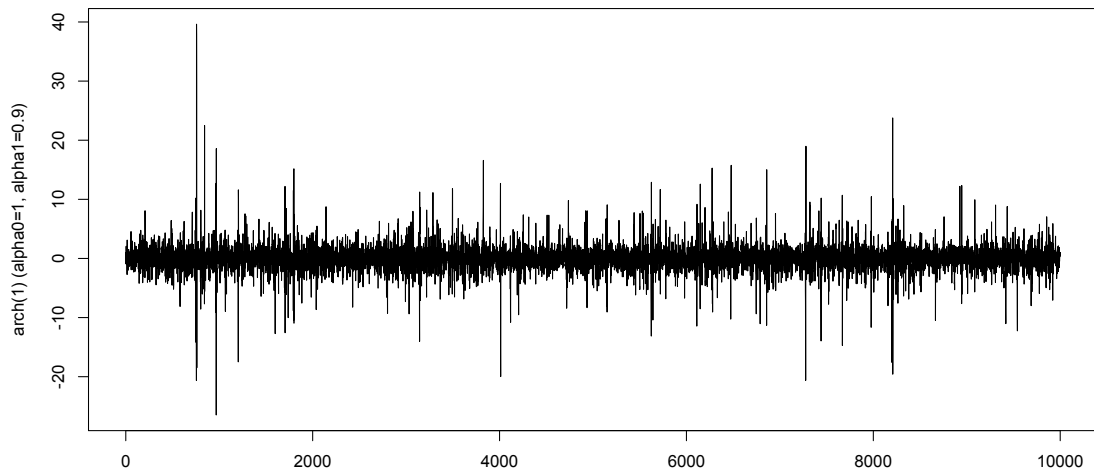
$$W_t = \beta W_{t-1} + \sqrt{\alpha_0 + \alpha_1 W_{t-1}^2} \varepsilon_t$$

erhält man eine ARCH(1)-Simulation in R durch

```
> w<-c(1:10000)
> alpha0<-0.5
> alpha1<-0.6
> w<-ts(w)
> for(i in 1:9999){
+ w[i+1]<-sqrt(alpha0 + alpha1*(w[i]^2))*eps[i+1]}
> arch1<-w
```

Mit zwei Parametern und der ARCH typischen Konstruktion, erhält man zum einen, eine erhöhte Variabilität, aber auch die besonderen Eigenschaften, wie etwa Clusterbildung.





Der Störterm ist wiederum identisch und der Parameter  $\alpha_0$  wird auf den Wert 1 festgelegt. Die erkennbaren Unterschiede in den obigen Grafiken/Simulationen werden also nur durch die Werte des Parameters  $\alpha_1 = 0.1, 0.5, 0.7$  und  $0.9$  bestimmt.

Mit diesen Prozess-Bausteinen sollte das Verständnis der Simulationsalgorithmen leichter fallen. Dem selben Prinzip folgend, werden diese wiederum auf den geglätteten Fall erweitert und bilden selbst auch nur eine Etappe für die Entwicklung von Parameterschätzungs- und Vorhersagealgorithmen (siehe Abschnitte 4.6, 4.7, 4.9).

## 4.6 garch11Fit

Neben der reinen Simulation von Prozessverläufen, ist man in der Praxis vor allem daran interessiert zu gegebenem Datenmaterial einen möglichst passenden Prozess zu finden. Die Anpassung erfolgt über die zur Verfügung stehenden Parameter. Theoretisch wäre natürlich mit immer mehr Parametern eine immer bessere Anpassung möglich. Um die Praktikabilität nicht zu gefährden, werden aber die Modellauswahl und auch die Anzahl der Parameter beschränkt.

Da der GARCH(1,1)-Prozess in der vorliegenden Arbeit keine entscheidende Rolle einnimmt, hätte der Algorithmus zur Parameterschätzung für diesen Prozess auch völlig entfallen können. Der `garch11Fit`-Algorithmus [22] dient jedoch als Ausgangspunkt für die selbstgeschriebenen Prozeduren `ar1arch1Fit`, `(Poly)Sar1arch1Fit` und sollte die gleichen Ergebnisse liefern, wie das weit verbreitete `garchFit`-Paket. Um die Funktionalität bzw. Korrektheit des `garch11Fit`-Algorithmus nachzuweisen, werden in [22] die Parameterschätzungen auf Grundlage des „`dem2gbp`“-Datensatzes<sup>20</sup> mit denen des `garchFit`-Algorithmus [3] verglichen. Die Ergebnisse dieser Funktionalitätsüberprüfung sind entscheidend für die Erwähnung an dieser Stelle und für einige Abwandlungen in Hinblick auf die Umsetzung auf die selbstgeschriebenen Algorithmen.

---

<sup>20</sup>Im R-Paket „fGarch“ enthaltener Datensatz des Wechselkurses von DM und Britischen Pfund - vgl. [22].

## Parameterschätzung für dem2gbp

Algorithmus: `garchFit`

$\alpha_0 = 0.010761384$ ,  $\alpha_1 = 0.153134058$ ,  $\beta = 0.805973748$

Algorithmus: `garch11Fit`

$\alpha_0 = 0.01076140$ ,  $\alpha_1 = 0.15313411$ ,  $\beta = 0.80597365$

Obwohl die Parameterschätzung in diesem Fall nahezu übereinstimmt, liefert der `garch11Fit`-Algorithmus, anders als in [22] propagiert, nicht immer die selben Ergebnisse wie `garchFit`. Als Beispiel dafür, können die Ergebnisse für den Allianz-Kurs herangezogen werden.

## Parameterschätzung für Allianz

Algorithmus: `garchFit`

$\alpha_0 = 0.000004357972$ ,  $\alpha_1 = 0.07764588$ ,  $\beta = 0.9124154$

Algorithmus: `garch11Fit`

$\alpha_0 = 0.0000000001432727$ ,  $\alpha_1 = 0.8139398$ ,  $\beta = 0.000001$

Das Problem scheint in der „Filter“-Repräsentation zu liegen, wie ein Hinweis im Programmcode von `garchFit` [3] nahe legt. Darüber hinaus wird für den Fall von Komplikationen bei den Berechnungen in `garchFit` so verfahren, dass auf eine ansonsten für Testzwecke vorbehaltene Version zurückgegriffen wird. Jene Testversion verzichtet auf die „Filter“-Repräsentation und wählt stattdessen eine Konstruktion über eine FOR-Schleife. Dies hat den Nachteil, dass die Berechnung wesentlich länger dauert (Minuten anstatt Sekunden), bietet aber den großen Vorteil, dass man ohne weiteres genau ablesen kann, was berechnet wird. Eine genauere Vorstellung über die Funktionsweise der Prozedur „Filter“, liefert [35] in Form von einigen Minimalbeispielen oder auch die Hilfe-Funktion von R.

Die Ergebnisse dieser vergleichsweise sehr langsamen FOR-Schleifen-Konstruktion lauten:

Algorithmus: `garch11Fit` - FOR-Schleife; Daten: dem2gbp

$\alpha_0 = 0.01076140$ ,  $\alpha_1 = 0.15313411$ ,  $\beta = 0.80597365$

Algorithmus: `garch11Fit` - FOR-Schleife; Daten: Allianz

$\alpha_0 = 0.00000435797$ ,  $\alpha_1 = 0.0776459$ ,  $\beta = 0.912415$

Die Unterschiede zu den Ergebnissen von `garchFit` sind nur noch minimal bzw. durch Rundung der letzten Dezimalstelle begründet und die längere Berechnungszeit spielt an dieser Stelle keine Rolle. Um die Vorteile der kurzen und überschaubaren Darstellung nicht aufgeben zu müssen, soll diese minimale Abweichung akzeptiert werden. Da der GARCH(1,1)-Fall in dieser Arbeit gar nicht behandelt wird, kann auf das nähere Studium des Programmcodes verzichtet werden. Der gesamte Abschnitt und die zwei Datensätze zeigen jedoch eindrucksvoll, dass man sich nicht blind auf etwaige Programmpakete bzw. deren Ergebnisse verlassen darf. Letztendlich kombiniert folgender Algorithmus Teile von `garch11Fit` [22] und `garchFit` [3].

```

> garch11FitFOR=function(x)
{
#Step 1: Initialize Time Series Globally:
...

#Step 2: Initialize Model Parameters and Bounds
...

#Step 3: Set Conditional Distribution Function:
...

#Step 4: Compose log-Likelihood Function:
garchLLH=function(parm){
mu=parm[1]; omega=parm[2]; alpha=parm[3]; beta=parm[4]
persistence=alpha+beta
N=length(x)
p=1
q=1
z=(x-mu); mvar=mean(z^2)
h=rep(omega+persistence*mvar,N)
for(i in 2:N){h[i]=omega+sum(alpha*(abs(z[i-(1:p)]))^2)+sum(beta*h[i-(1:q)])}
hh=sqrt(abs(h))
llh=-sum(log(garchDist(z, hh)))
llh}
print(garchLLH(params))

#Step 5: Estimate Parameters and Compute Numerically Hessian:
...

#Step 6: Create and Print Summary Report:
...
}

```

## 4.7 Erläuterung ar1arch1Fit

Auch wenn der Programmcode zu `ar1arch1Fit` bzw. `Sar1arch1Fit` sich auf das Nötigste beschränkt (vgl. Abschnitt 3.4), ist das sofortige Verständnis keineswegs garantiert. Die folgenden Zeilen sind der Versuch, die einzelnen Schritte in aller Kürze verbal zu umreißen und so dem interessierten Leser einen schnelleren Einblick zu ermöglichen:

In **Step 1** und **Step 2** werden die Ursprungsdaten global gesichert, sowie die Startwerte und Schranken für die Parameterschätzung daraus abgeleitet. Besonders bei der Festlegung der Startwerte und Schranken gibt es kein unumstößliches Rezept, sondern eine gewisse Willkür. Hierbei sind die vorliegenden Werte so gewählt, dass sie die typischen Anwendungssituationen erfüllen - vgl. [22].



Da der Störterm jeweils als standardnormalverteilt angenommen wird, kann man für die bedingte Verteilungsfunktion ebenfalls eine Normalverteilung annehmen. Der Konstruktion in **Step 3** liegt die Dichte der Standardnormalverteilung zu Grunde (**dnorm**) und der erste Aufruf von **arlarch1Dist** erfolgt in **Step 4**. Hier zwar nicht gebraucht, aber vielleicht erwähnenswert: (**pnorm**) entspräche der Verteilungsfunktion.

Analog zu **garch11Fit** wird in **Step 4** die Log-Likelihood-Funktion umgesetzt. Wegen des AR-Parts wird bereits, für die Berechnung von **z**, eine FOR-Schleife verwendet. Bemerkenswert ist darüber hinaus, dass die Angabe der Parameter **p**, **q** bzw. hier **u**, **p** nur verdeutlichen soll, dass eine Erweiterung auf andere Parameterkonstellationen problemlos möglich wäre. Man könnte auch darauf verzichten, diese extra kenntlich zu machen, jedoch würde das den Programmcode nur unwesentlich verschlanken und zudem den Einblick darin verhindern, wo der jeweilige Parameter zum tragen kommt. Die Variable **h** spiegelt den auf Grund der Daten geschätzten Verlauf von  $\sigma^2$  wider.

Erst in **Step 5** erfolgt die eigentliche Parameterschätzung durch das in R enthaltene nichtlineare Optimierungsverfahren **nlmminb**. Dabei wird über **arlarch1LLH** optimiert und zwar mit den oben definierten Start- und Grenzwerten. Darüber hinaus wird die Hesse-Matrix numerisch berechnet, um eine Fehlerabschätzung vornehmen zu können. Die Hesse-Matrix enthält als Einträge die zweifachen partiellen Ableitungen und stellt bei 4 zu schätzenden Parametern eine  $4 \times 4$  - Matrix dar, um alle möglichen Kombinationen abzudecken. Da die genaue Berechnungsvorschrift der Einträge etwas befremdlich wirken kann, soll sie hier kurz motiviert werden.

Die erste partielle Ableitung in  $i$ -Richtung oder gemäß der  $i$ -ten Variablen, lässt sich darstellen in der Art:

$$\begin{aligned} \lim_{\varepsilon_i \rightarrow 0} \frac{f(x + \varepsilon_i) - f(x - \varepsilon_i)}{2\varepsilon_i} &= \lim_{\varepsilon_i \rightarrow 0} \frac{f(x + \varepsilon_i) - f(x) + f(x) - f(x - \varepsilon_i)}{2\varepsilon_i} \\ &= \lim_{\varepsilon_i \rightarrow 0} \frac{1}{2} \cdot \frac{f(x + \varepsilon_i) - f(x)}{\varepsilon_i} + \lim_{\varepsilon_i \rightarrow 0} \frac{1}{2} \cdot \frac{f(x) - f(x - \varepsilon_i)}{\varepsilon_i} = \frac{1}{2} \cdot f'_i(x) + \frac{1}{2} \cdot f'_i(x) = f'_i(x) \end{aligned}$$

Leitet man diese nun noch in  $j$ -Richtung ab, so ergibt sich:

$$f''_{ij}(x) = \lim_{\varepsilon_i \rightarrow 0, \varepsilon_j \rightarrow 0} \frac{(f(x + \varepsilon_i + \varepsilon_j) - f(x - \varepsilon_i + \varepsilon_j)) - (f(x + \varepsilon_i - \varepsilon_j) - f(x - \varepsilon_i - \varepsilon_j))}{2\varepsilon_i 2\varepsilon_j}$$

und spiegelt damit genau die Konstruktion in **Step 5** wider.

**Step 6** gibt die gesammelten Ergebnisse aus und berechnet zu den „geschätzten Parameterwerten“ weitere Kennzahlen, wie etwa Standard Errors - vgl. [22] S.30, die ein Maß für die Güte der Parameterschätzung darstellen. Die Standard Errors sollten von der Größenordnung kleiner sein als die berechneten Parameter und sind untereinander erst einmal nicht vergleichbar. Um eine Vergleichbarkeit zu gewährleisten, werden zusätzlich die **t-values** angegeben, die sich als Verhältnis von Parameterschätzung und Standard Error ergeben.

## 4.8 Residuenanalyse Smooth-AR(1)-ARCH(1)

Das folgende Programm errechnet im Smooth-AR(1)-ARCH(1)-Fall die Residuen. Die notwendigen Parameter ergeben sich aus der Anwendung von `Sar1arch1Fit`. Die Berechnungsschritte folgen einfach der Definition der Residuen.

Residuen `Sar1arch1`

```
x<-Allianz
y<-rep(0,3909)
beta<-0.138581
a0<-0.000279295
a1<-0.404023
alpha<-0.5
gamma<-0.5
w0unten<- -0.15
w0oben<-0.12
for(i in 1:3909){
  if(mu+beta*x[i+1]>w0oben){
    y[i+1]<-(x[i+1]-w0oben+gamma*w0oben-gamma*mu-gamma*beta*x[i])}
  if(mu+beta*x[i+1]<w0unten){
    y[i+1]<-(x[i+1]-w0unten+alpha*w0unten-alpha*mu-alpha*beta*x[i])}
  if(mu+beta*x[i+1]>w0unten && mu+beta*x[i+1]<w0oben){
    y[i+1]<-(x[i+1]-mu-beta*x[i])
  }
}
```

Gemäß der Aufteilung unterhalb, innerhalb und oberhalb des Kompaktums, ergibt sich jeweils eine andere Berechnungsvorschrift und macht diese Abfolge von `if`-Abfragen nötig. Im AR(1)-ARCH(1)-Fall vereinfacht sich dies zu:

```
for(i in 1:3909){
  y[i+1]<-(x[i+1]-mu-beta*x[i])}.
```

## 4.9 Predict-Poly-Smooth

Folgender Algorithmus liefert eine Vorhersage/Schätzung der nächsten `n.ahead` Datenpunkte und gründet dieses Ergebnis auf den Median von jeweils `n.sim` Simulationen. Darüber hinaus kann über die Art der Glättungsfunktion ("`smooth`", "`poly`"), je nach Anwendungssituation entschieden werden. Letztendlich handelt es sich hierbei um eine Erweiterung des Minimalprogramms in Abschnitt 3.4.2. und es soll auf diese Weise dem allgemeinen Fall mit den möglichen Konstellationen Genüge getan werden.

```

PredictPolySmoothar1arch1=function(x,g,mu,beta,alpha0,alpha1,alpha,gamma,
n.ahead,n.sim,w0oben,w0unten){
z=rep(0,(n.ahead+1))
z[1]=x[length(x)]
y=rep(0,n.sim)
if(g=="smooth") { for(i in 2:n.ahead) {
set.seed(i)
eps<-rnorm(n.sim)
for(j in 1:n.sim) {
y[j]=mu+beta*z[i]+sqrt(alpha0+alpha1*z[i]^2)*eps[j]
if(y[j]>w0oben)y[j]<-w0oben+gamma*(y[j]-w0oben)
if(y[j]<w0unten)y[j]<-w0unten+alpha*(y[j]-w0unten)
y[j]<-y[j]
}
z[i+1]=median(y)
}
} if(g=="poly") { for(i in 1:n.ahead) {
set.seed(i)
eps<-rnorm(n.sim)
for(j in 1:n.sim) {
y[j]=mu+beta*z[i]+sqrt(alpha0+alpha1*z[i]^2)*eps[j]
if(y[j]>w0oben)y[j]<-(w0oben-1)+(y[j]-w0oben+1)^gamma
if(y[j]<w0unten)y[j]<-(w0unten+1)+(w0unten+1-y[j])^alpha
y[j]<-y[j]
}
z[i+1]=median(y)
}
}
z[2:(n.ahead+1)]
}

```

In der obigen Version muss man die Ergebnisse der Parameterschätzung beim Aufruf noch selbst eingeben. Für dauerhaften Gebrauch erscheint das wenig sinnvoll und es wäre sicherlich ratsam die bisherigen Prozeduren so zu verknüpfen, dass man die Parameter beim Aufruf automatisch ausliest. An dieser Stelle bringt dies allerdings keinen nennenswerten Mehrwert.

`n.ahead=4` und `n.sim=1000` liefern beispielsweise (in Übereinstimmung mit dem Minimalprogramm):

	Predict
1	-0.0001336027
2	0.0005616702
3	0.0003607027
4	-0.0008728005

Da alle Ergebnisse im Kompaktum  $[-0.15, 0.12]$  liegen, unterscheiden sich die Vorhersagen für die unterschiedlichen Glättungsfunktionen nicht. Ändert man jedoch den letzten `x`-Datenpunkt

(Startpunkt für die Vorhersage), so kann man sehr schnell zu unterschiedlichen Werten gelangen.

Der entscheidende Unterschied zu der Vorgehensweise im ungeglättet Fall, liegt in der Zufallsabhängigkeit. Während im ungeglätteten Fall der Zufall unter den Tisch viel und sich immer nur genau eine Vorhersage anbot, sind die Ergebnisse im geglätteten Fall sehr wohl zufallsabhängig. So führt ein unterschiedlicher Simulationsumfang (`n.sim`) zu verschiedenen Werten. Auch wenn in der Berechnungsvorschrift selbst kein Zufall zum tragen kommt, so ist er doch Modellbedingt immer vorhanden und macht insbesondere weitreichende Vorhersagen mehr und mehr fragwürdig. Genau wie im ungeglätteten Fall auch, ist es äußerst wichtig sich die Entstehungsweise deutlich zu machen, bevor man blind auf eine Vorhersage vertraut.

## 4.10 Normalverteilung und Errorfunktion - Fehlerfunktion

Die Normalverteilung oder besser gesagt die Standardnormalverteilung ist weit verbreitet und ihre Verteilungsfunktion in Form von tabellierten Werten bzw. durch Einbindung in Standardprogramme jederzeit verfügbar<sup>21</sup>.

In diesem Zusammenhang findet man auch immer wieder die Gaußsche Fehlerfunktion, da Sie nicht nur eine gewisse Ähnlichkeit zur Verteilungsfunktion der Normalverteilung hat, sondern sich mit dieser in Beziehung setzen lässt. Wie der Name schon verrät, ist ein möglicher Anwendungsbereich die Fehlerabschätzung und zu diesem Zwecke soll sie hier auch dienen.

Dazu sind einige Eigenschaften und Zusammenhänge von Nöten, die hier gesammelt notiert werden sollen - vgl. [1].

Die Definition der Fehlerfunktion

$$\operatorname{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x e^{-\tau} d\tau$$

und deren Komplementären-Funktion

$$\operatorname{erfc}(x) = 1 - \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-\tau} d\tau$$

stellen bereits die Verbindung zur Verteilungsfunktion der Normalverteilung heraus. Sie haben beide einen Wertebereich von -1 bis 1.

Für unsere Zwecke ist aber weniger diese Definition von Nutzen, als deren Implikationen.

Ein entscheidender Punkt ist die Tatsache, dass man es mit einer ungeraden Funktion zu tun hat

$$\operatorname{erf}(-x) = -\operatorname{erf}(x).$$

Des Weiteren ist der konkrete Zusammenhang zur Verteilungsfunktion der Standardnormalverteilung gegeben durch

$$\Phi(x) = \frac{1}{2} \left( 1 + \operatorname{erf} \left( \frac{x}{\sqrt{2}} \right) \right)$$

---

<sup>21</sup>vgl. `pnorm` in R

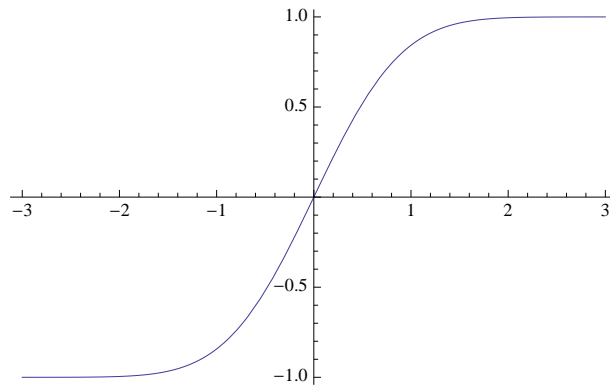


Abbildung 4.1: Kurvenverlauf Errorfunktion

und im Falle einer Normalverteilung

$$F(x) = \frac{1}{2} \left( 1 + \text{erf} \left( \frac{x - \mu}{\sigma\sqrt{2}} \right) \right).$$

Die enge Beziehung zur Exponentialverteilung und somit eine Möglichkeit zur besseren Abschätzung, erkennt man am besten an den jeweiligen Reihendarstellungen:

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)n!} = \frac{2}{\sqrt{\pi}} \left( x - \frac{x^3}{3} + \frac{x^5}{10} - \frac{x^7}{42} + \frac{x^9}{216} - \dots \right)$$

und

$$e^{-z} = \sum_{n=0}^{\infty} \frac{(-1)^n z^n}{n!}.$$

In vielen Situationen kann man obige Zusammenhänge nutzen, um eine schnelle und doch genügend scharfe Abschätzung zu erhalten. Im Zusammenhang mit dem Verzicht auf die Jensensche Ungleichung, führte sie im polynomialen Ansatz jedoch nicht zu dem erhofften Ergebnis. Nichts desto trotz hat man so ein weiteres nützliches Werkzeug zur Hand und verschafft sich einen tieferen Einblick.

## 4.11 Streng Stationärer Innovationsprozess

Die unterschiedliche Ausgangssituation im Falle eines i.i.d. oder eines streng stationären Innovationsprozesses wurde bereits zu Beginn angesprochen. Welche Veränderungen eine Abkehr von der i.i.d. Situation im einzelnen zu Folge hat und wie man trotzdem zu ähnlichen Aussagen gelangen kann, soll in diesem Abschnitt zumindest angedeutet werden.

Anders als im i.i.d. Fall, hat man es nun nicht mehr mit einer Markov Kette zu tun und muss sich eines vollkommen anderen Instrumentariums bedienen. Da man sich somit ein neues Schlachtfeld eröffnen würde, soll sich dieser Abschnitt darauf beschränken den zentralen Beweis zu Chapter 6 [16] etwas zugänglicher zu machen. So ergibt sich zumindest ein kurzer Einblick in die auf den ersten Blick nur minimal, aber doch folgeschwer veränderte Situation.

Neben der Forderung nach einem streng stationären Innovationsprozesses mit unendlicher Vergangenheit (vgl. Prop. 4.6 - [16]), werden im Rahmen des IFS die Übergangsfunktionen  $u_x : W \rightarrow W$ ,  $x \in X$  als Lipschitz angenommen. Für die Beweisführung wird zunächst in Analogie zum i.i.d. Fall

$$\zeta_n^w = u_{X_n} \circ \cdots \circ u_{X_1}(w), \quad w \in W$$

für  $n \in \mathbb{N}$  definiert. Hierbei beschreibt  $\zeta_0^w = w$  einen Startwert und  $(\zeta_n^w)_{n \in \mathbb{N}}$  einen stochastischen Prozess, allerdings ohne Markov-Eigenschaft.

### Satz 4.1. (*Proposition 6.1 - [16]*)

Seien die  $u_x$ ,  $x \in X$  Lipschitz-Funktionen mit

$$\int_{\Omega} \log^+ s(u_{X_0}) dP < \infty.$$

Dann gilt:

- (i) Es existiert eine invariante Funktion (Lyapunov-Exponent)  $\chi : \Omega \rightarrow \mathbb{R} \cup \{-\infty\}$  und  $\chi^+ \in L^1(P)$ . Weiter gilt:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log s(\zeta_n^\bullet) = \chi \quad Pf.s.$$

und

$$\lim_{n \rightarrow \infty} \int_{\Omega} \frac{1}{n} \log s(\zeta_n^\bullet) dP = \inf_{n \in \mathbb{N}} \frac{1}{n} \int_{\Omega} \log s(\zeta_n^\bullet) dP = \int_{\Omega} \chi dP.$$

- (ii) Für alle  $k \in \mathbb{Z}$  gilt:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log s(u_{X_{k-1}} \circ \cdots \circ u_{X_{k-n}}) = \chi \quad Pf.s.$$

Der Lyapunov-Exponent wurde oben bereits als Kriterium eingeführt ( $\chi < 0$ ) und findet auch im streng stationären Fall seine Anwendung.

**Satz 4.2. (Theorem 6.2 - [16])**

Seien die  $u_x$ ,  $x \in X$  Lipschitz-Funktionen mit

$$\int_{\Omega} \log^+ s(u_{X_0}) dP < \infty$$

und

$$\int_{\Omega} \log^+ d(w_0, u_{X_0}(w_0)) dP < \infty$$

für ein  $w_0 \in W$ . Weiter sei  $\chi < 0$  P.f.s. (siehe Satz 3.1.).

(i) Es existiert

$$\eta_k = \lim_{m \rightarrow \infty} u_{X_k} \circ \cdots \circ u_{X_{k-m+1}}(w)$$

für alle  $w \in W$ ,  $k \in \mathbb{Z}$  und hängt nicht von der Wahl von  $w$  ab. Der stochastische Prozess  $(\eta_k, k \in \mathbb{Z})$  nimmt Werte in  $W$  an und ist streng stationär. (Da  $\eta_n = u_{X_n} \circ \cdots \circ u_{X_1} \circ \eta_0$  P.f.s., ist  $\eta_0$  eine Zufallsvariable von Startwerten, die  $(\zeta_n^w)_{n \in \mathbb{N}}$  streng stationär werden lassen.)

(ii) Für alle  $w \in W$  konvergiert  $(\zeta_n^w, \zeta_{n+1}^w, \dots)$  in Verteilung gegen  $(\eta_0, \eta_1, \dots)$ , für  $n \rightarrow \infty$ . Insbesondere konvergiert also  $\zeta_n^w$  in Verteilung gegen  $\eta_0$ , für  $n \rightarrow \infty$ .

Im Original umfasst der obige Satz noch weitere Aussagen, die für uns allerdings nicht von Bedeutung sind. Den Beweis findet man zwar in Iosifescu [16], jedoch in so kompakter Form, dass ein direktes lesen bzw. verstehen äußerst schwer erscheint. Das Folgende ist demnach der Versuch, einige Gedankensprünge nachzuvollziehen.

**Beweis zu Satz 4.2.**

(i) Zunächst ergibt sich:

$$\mathbb{E}(\log^+ d(w, u_{X_0}(w))) = \int_{\Omega} \underbrace{\log^+ d(w, u_{X_0}(w))}_{(***)} dP < \infty$$

auch für alle  $w \in W$ , denn

$$\begin{aligned} (**) &\leq \log^+(d(w, w_0) + d(w_0, u_{X_0}(w_0)) + d(u_{X_0}(w_0), u_{X_0}(w))) \\ &\leq \max \left( \log^+ d(w_0, u_{X_0}(w_0)), \log^+ \left( d(w, w_0) + \underbrace{d(u_{X_0}(w_0), u_{X_0}(w))}_{\leq d(w_0, w) \text{ da Lipschitz}} \right) \right) + \log 2 \end{aligned}$$

und somit das Geforderte durch die Voraussetzung  $\int_{\Omega} \log^+ d(w_0, u_{X_0}(w_0)) dP < \infty$ . Die letzte Ungleichung erklärt sich durch:

$$\log^+(a + b) \leq \max(\log^+ a, \log^+ b) + \log 2, \quad \text{für positive } a \text{ und } b.$$

Für jedes beliebige, aber dann fixe  $k \in \mathbb{Z}$ , sei nun

$$\zeta_{k,m} := u_{X_k} \circ u_{X_{k-1}} \circ \cdots \circ u_{X_{k-m+1}}, \quad \text{für } m \in \mathbb{N}_+$$

und damit

$$d(\zeta_{k,m}^w, \zeta_{k,m+1}^w) = \frac{d(\zeta_{k,m}^w, \zeta_{k,m}^{u_{X_{k-m}}(w)})}{d(w, u_{X_{k-m}}(w))} \cdot d(w, u_{X_{k-m}}(w)) \leq \underbrace{s(\zeta_{k,m}^\bullet)}_{(*)} \underbrace{d(w, u_{X_{k-m}}(w))}_{(**)}$$

für alle  $w \in W$ . Sei  $\Omega_j = \{\chi < -1/j\}$ ,  $j \in \mathbb{N}_+$ , so dass  $P(\cup_{j \in \mathbb{N}_+} \Omega_j) = 1$ , da ja  $\chi < 0$  nach Voraussetzung.

Mit Satz 3.1 (ii), ergibt sich für alle  $\omega \in \Omega_j$  und  $m$  hinreichend groß (in Abhängigkeit von  $\omega$ ):

$$\frac{1}{m} \log s(\zeta_{k,m}^\bullet) < -1/j$$

bzw.

$$s(\zeta_{k,m}^\bullet) < \exp(-m/j).$$

Damit erhält man also eine Abschätzung für (\*).

Für alle  $j \in \mathbb{N}$  und  $w \in W$  gilt

$$\begin{aligned} \sum_{m \in \mathbb{N}} P(\log^+ d(w, u_{X_{k-m}}(w)) > m/2j) &\stackrel{(*4)}{=} \sum_{m \in \mathbb{N}} P(\log^+ d(w, u_{X_0}(w)) > m/2j) \\ &\stackrel{(*5)}{\leq} 2j \mathbb{E}(\log^+ d(w, u_{X_0}(w))) \end{aligned}$$

Hierbei wird in (\*4) die strenge Stationarität von  $(X_n, n \in \mathbb{Z})$  ausgenutzt, da in diesem Zusammenhang nur identisch verteilt, aber nicht unabhängig wichtig ist.

Die Ungleichung (\*5) ergibt sich aus dem Zusammenhang  $\mathbb{E}[Y] = \int_0^\infty 1 - F(y) dy$  für eine Z.V.  $Y \geq 0$  und einer Diskretisierung des Integrals:

$$\begin{aligned} \mathbb{E}(\log^+ d(w, u_{X_0}(w))) &= \int_0^\infty P(\log^+ d(w, u_{X_0}(w)) > y) dy = \sum_{m \in \mathbb{N}} \int_{(m-1)/2j}^{m/2j} P(\log^+ d(w, u_{X_0}(w)) > y) dy \\ &\geq \frac{1}{2j} \sum_{m \in \mathbb{N}} P(\log^+ d(w, u_{X_0}(w)) > m/2j). \end{aligned}$$

Da ja  $\mathbb{E}(\log^+ d(w, u_{X_0}(w))) < \infty$  (siehe Anfang des Beweises), hat man dank (\*4) und (\*5) eine Konstellation wie im Borell-Cantelli Lemma. Im klassischen Fall, hängt das  $\varepsilon$  zwar nicht von  $m$  ab, aber es ist durchaus zulässig - vgl. Theorem 1.5.1 [4] für diese und weitere Abwandlungen bzw. Anwendungsbeispiele.



Formal wird die Aussage

$$\sum_{m \geq 1} P(A_m) < \infty \Rightarrow P(A) = 0, \quad \text{mit} \quad A = \limsup A_m = \{A_m \text{ unendlich oft}\}$$

und  $(\limsup A_m)^C = \liminf A_m^C$  verwendet, um zu schlussfolgern

$$\begin{aligned} P(\limsup A_m) &= P(\limsup (\log^+ d(w, u_{X_{k-m}}(w)) > m/2j)) = P(\log^+ d(w, u_{X_{k-m}}(w)) > m/2j \text{ i.o.}) = 0 \\ &\Rightarrow P((\limsup A_m)^C) = P(\liminf A_m^C) = P(\log^+ d(w, u_{X_{k-m}}(w)) \leq m/2j \text{ eventually}) = 1. \end{aligned}$$

Das heißt wiederum nichts anderes als

$$d(w, u_{X_{k-m}}(w)) \leq \exp(m/2j)$$

ab einem hinreichend großen  $m$ , für fast alle  $\omega \in \Omega_j$  und somit eine Abschätzung für (\*\*). Setzt man nun die Abschätzungen für (\*) und (\*\*) ein, so ergibt sich

$$d(\zeta_{k,m}^w, \zeta_{k,m+1}^w) \leq \exp(-m/2j), \quad w \in W$$

für  $m$  hinreichend groß und fast alle  $\omega \in \Omega_j$ , so dass  $(\zeta_{k,m})_{m \in \mathbb{N}_+}$  eine Cauchy-Folge in  $\Omega_j$  ist. Diese konvergiere gegen die Zufallsvariable  $\eta_k$ .

Abschliessend erhält man

$$d(\zeta_{k,m}^{w'}, \zeta_{k,m}^{w''}) = \frac{d(\zeta_{k,m}^{w'}, \zeta_{k,m}^{w''})}{d(w', w'')} \cdot d(w', w'') \leq \sup_{w', w''} \frac{|\zeta_{k,m}^{w'} - \zeta_{k,m}^{w''}|}{d(w', w'')} \cdot d(w', w'') = s(\zeta_{k,m}^\bullet) \cdot d(w', w'')$$

für alle  $w', w'' \in W$  und  $m \rightarrow \infty$ , was die Unabhängigkeit von  $w$  zeigt. Die strenge Stationarität von  $(\eta_k, k \in \mathbb{Z})$  folgt aus der für  $(X_k, k \in \mathbb{Z})$  ( $X_i \stackrel{\nu}{=} X_j$  überträgt sich auf  $u_{X_i} \stackrel{\nu}{=} u_{X_j}$ ).

(ii) Da  $\zeta_n^w = \zeta_{n,n}^w$  für  $n \in \mathbb{N}$ , ergibt sich gemäß Stationarität auch, dass  $(\zeta_n^w, \zeta_{n+1}^w, \dots)$  die selbe Verteilung besitzt wie  $(\zeta_{0,n}^w, \zeta_{1,n+1}^w, \dots)$  und nach (i) gegen  $(\eta_0, \eta_1, \dots)$  konvergiert für  $n \rightarrow \infty$  (Verschiebung von  $n+1$  auf 1, gemäß Stationarität).  $\square$

Allein der obige Beweis macht deutlich wie umfangreich die notwendigen Anpassungen sind, um im Fall eines streng stationären Innovationsprozesses zu vergleichbaren Ergebnissen zu gelangen. Alle bisherigen Überlegungen müsste man konsequenterweise auf diese neue Ausgangssituation hin anpassen, um sich so Schritt für Schritt der Konsequenzen bewusst zu werden. Mit zunehmendem Praxisbezug wird der Verzicht auf starke Forderungen wie i.i.d. erforderlich werden und bietet also noch ein weites Betätigungsfeld für die Untersuchung von Glättungsfunktionen, jedoch unter deutlich veränderten Rahmenbedingungen.

Neben dieser grundlegend neuen Ausrichtung für weitere Untersuchungen, stehen vor dem Hintergrund der gemachten Ergebnisse aber auch im i.i.d. Fall noch einige Erweiterungsmöglichkeiten im Raum. So zog sich das AR(1)-ARCH(1)-Modell wie ein roter Faden durch die gesamte Arbeit und bot an vielen Stellen die konkrete Grundlage, um die Konsequenzen durch

Glättungsfunktionen zu untersuchen. Nachdem die Grundlagenarbeit aber getan ist und viele Ergebnisse sich als allgemeingültig erwiesen haben, würde also die konkrete Anwendung auf verwandte autoregressive Prozesse oder sogar noch exotischere Exemplare genügend Spielraum für weitere Untersuchungen bieten. Eine praktische Anwendung des Glättungsansatzes in weiteren Situationen steht also noch aus, bevor endgültig über seinen Wert entschieden werden kann.

## Ausblick

In dieser Arbeit wurden einige Themen behandelt bzw. Probleme angesprochen, die eine weitere Untersuchung verdienen.

So sollte der Einfluss „aufgesetzter“ Glättungsfunktionen auf die Momente und Endstücke (= tails) der stationären Verteilung weiter untersucht werden. Zu erwarten ist dabei, dass die stationäre Grenzverteilung „gutartiger oder gezügelter“ wird, hinsichtlich der Existenz von Momenten und, damit zusammenhängend, der Besetzung ihrer Endstücke.

Die hier durchgeführten empirischen Untersuchungen zeigen, dass der Einsatz eines AR(1)-ARCH(1)-Modells zur Modellierung von Aktienkursverläufen keinesfalls unkritisch zu sehen ist. Auch, oder gerade, wenn klar ist, dass hier niemand „den Stein der Weisen“ in Besitz hat, sollte die Suche nach geeigneten Modellen weiter gehen. Innerhalb der Klasse der AR(1)-ARCH(1)-Modelle verdienen die Modifikationen mit Glättungsfunktionen weitere Aufmerksamkeit.

Schließlich bietet die in 4.11 vorgestellte Modellbildung eines streng stationären Innovationsprozesses unter theoretischen wie auch empirischen Aspekten ein weites Betätigungsfeld. Die zitierte Arbeit von Iosifescu [16] gibt dazu gerade den Einstieg her. Da in diesem Fall der interessierende Prozess nicht mehr die Markov-Eigenschaft besitzt, sind zu seiner Untersuchung andere theoretische Methoden bzw. Hilfsmittel heranzuziehen, d.h. man muss eine neues „Spiel-feld“ betreten.

## Literatur

- [1] N.L. Johnson & S. Kotz & N. Balakrishnan. *Continuous univariate distributions. 2 ed.* Wiley, New York, 1994.
- [2] B. Basrak. *The Sample Autocorrelation Function of Non-Linear Time Series*. PhD thesis, Rijksuniversiteit Groningen, 2000.
- [3] D. Würtz & Y. Chalabi. *Package “fGarch” - garchFit*. R-Project, <https://svn.r-project.org/Rmetrics/trunk/fGarch/R/garchFit.R>, 2012. <http://cran.r-project.org/web/packages/fGarch/fGarch.pdf>.
- [4] T.K. Chandra. *The Borel-Cantelli Lemma*. Springer, India, 2012.
- [5] D.B.H. Cline. Regular variation of order 1 nonlinear AR-ARCH models. *Stochastic Processes and their Application*, 117:840–861, 2007.
- [6] P.J. Brockwell & R.A. Davis. *Introduction to Time Series and Forecasting*. Springer, New York, 1996.
- [7] P. Diaconis & D. Freedman. Iterated Random Functions. *SIAM Rev.*, 41(1):45–76, 1999.
- [8] A.J. McNeil & R. Frey. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7:271–300, 2000.
- [9] H.-O. Georgii. *Stochastik 4. Auflage*. de Gruyter, Berlin, 2009.
- [10] J. Groß. *Grundlegende Statistik mit R*. Vieweg+Teubner, Wiesbaden, 2010.
- [11] J. Diebolt & D. Guégan. Tail behavior of the stationary density of general non-linear autoregressive processes of order 1. *Journal of Applied Probability*, 30:315–329, 1993.
- [12] J.R. Harvey. *Fractional Moments of a Quadratic Form in Noncentral Normal Random Variables*. North Carolina State University at Raleigh, 1965.
- [13] U. Herkenrath. Autoregressive Time Series Models with Smoothed Extreme Values. *Advances in Mathematics, Invited Contributions to the seventh Congress of Romanian Mathematicians*, 2013.
- [14] T. Hothorn. *Vorlesungsskript Statistik III*. Institut für Statistik, Ludwig-Maximilians-Universität München, 2008.
- [15] F. Hoyer. Eine Alternative zur Korrelationsfunktion - Vergleich verschiedener nicht-linearer Modelle. Master’s thesis, Technische Universität München, 2006.
- [16] M. Iosifescu. Iterated function systems: A critical survey. *Schriftenreihe des Fachbereichs Mathematik*, SM-DU-692, 2009.
- [17] N. Kiefersbeck. Stationarität und Tailindex von zeitdiskreten Volatilitätsmodellen. Master’s thesis, Technische Universität München, 1999.
- [18] A. Klenke. *Wahrscheinlichkeitstheorie*. Springer, Berlin, 2006.

- [19] M. Borkovec & C. Klüppelberg. The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. *The Annals of Applied Probability*, Vol. 11(4):1220–1241, 2001.
- [20] M. Kolonko. *Stochastische Simulation*. Vieweg+Teubner, Wiesbaden, 2008.
- [21] G. Letac. a contraction principle for certain Markov chains and its applications. *Contemporary Mathematics*, Vol. 50:263–273, 1986.
- [22] D. Würtz & Y. Chalabi & L. Luksan. Parameter Estimation of ARMA Models with GARCH/APARCH Errors - An R and SPlus Software Implementation. *Journal of Statistical Software*, forthcoming.
- [23] G. Maercker. *Statistical inference in conditional heteroskedastic autoregressive models*. PhD thesis, Universität Braunschweig, 1997.
- [24] C. Klüppelberg & L. Peng. Empirical Likelihood Methods for an AR(1) Process with ARCH(1) Errors. Sonderforschungsbereich 386 Paper 469, Ludwig-Maximilian-Universität München, 2006.
- [25] P. Bougerol & N. Picard. Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics*, Vol. 52:115–127, 1992.
- [26] A. Rudolph. Stochastic linear autoregressive models. *working paper*, 2013.
- [27] R. Schlittgen. *Einführung in die Statistik*. Oldenbourg, München, 1997.
- [28] D. Straumann. *Estimation in Conditionally Heteroscedastic Time Series Models*. Springer, Berlin, 2005. ISBN 3-540-21135-7.
- [29] R. Schlittgen & B. Streitberg. *Zeitreihenanalyse*. Oldenbourg, München, 2001.
- [30] L. Truquet. A new smoothed QMLE for AR process with LARCH errors. <http://hal.archives-ouvertes.fr/docs/00/28/48/84/PDF/larch0306.pdf>, 2008.
- [31] R.S. Tsay. *Analysis of Financial Time Series - Second Edition*. Wiley-Interscience, Hoboken NJ, 2005.
- [32] P.D. Feigin & R.L. Tweedie. Random coefficient autoregressive process: a Markov chain analysis of stationarity and finiteness of moments. *Journal of Time Series Analysis*, Vol. 6:1–14, 1985.
- [33] R.L. Tweedie. The existence of moments for stationary Markov chains. *Journal of Applied Probability*, Vol. 20:191–196, 1983.
- [34] S.P. Meyn & R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer, Berlin, 1993.
- [35] A.W. van der Vaart. Time Series in S-Plus. *Faculteit Wiskunde en Informatica*, Vrije Universiteit: <http://www.math.vu.nl/aad/timeseries/splus.eps>, 1998.

- [36] H. Wang. *Multivariate Dichteschätzung in der explorativen Datenanalyse*. PhD thesis, Universität Augsburg, 2009.
- [37] J. Fan & Q. Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, Berlin, 2002.
- [38] C. Francq & J.-M. Zakoian. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bdennoulli*, 10(4):605–637, 2004.
- [39] W. Zucchini. *Vorlesungsskript Zeitreihenanalyse*. Universität Göttingen, online unter: [http:// www.stat.oek.wiso.uni-goettingen.de/veranstaltungen/zeitreihen/sommer04/](http://www.stat.oek.wiso.uni-goettingen.de/veranstaltungen/zeitreihen/sommer04/), 2004.